

UNIVERSIDAD AUTÓNOMA DE CHIHUAHUA
FACULTAD DE FILOSOFÍA Y LETRAS
SECRETARÍA DE INVESTIGACIÓN Y POSGRADO



**UNIVERSIDAD AUTÓNOMA DE
CHIHUAHUA**

**EL LECTOR MODELO DE CHATGPT: ANÁLISIS DE LA COOPERACIÓN
INTERPRETATIVA EN LAS CREACIONES LITERARIAS DE LAS INTELIGENCIAS
ARTIFICIALES GENERATIVAS**

POR:

TOMÁS EDUARDO TENA ACOSTA

**TESIS PRESENTADA COMO REQUISITO PARA OBTENER EL GRADO DE
MAESTRO EN INVESTIGACIÓN HUMANÍSTICA**

NOVIEMBRE 2025



Tena 1



CHIHUAHUA, CHIH. MÉXICO

NOVIEMBRE DEL 2025

"El Lector Modelo de ChatGPT: análisis de la cooperación interpretativa en las creaciones literarias de las inteligencias artificiales generativas". Tesis presentada por Tomás Eduardo Tena Acosta como requisito parcial para obtener el grado de Maestro en Investigación Humanística ha sido aprobado y aceptado por:

Dr. Javier Horacio Contreras Orozco
Director de la Facultad de Filosofía y Letras

Dr. Jorge Alan Flores Flores
Secretario de Investigación y Posgrado

M. H. Victor Manuel Córdova Pereira
Coordinador Académica

Dr. Arturo Rico Bovio
Presidente

Fecha: noviembre del 2025 Comité:

Director(a) de Tesis: Dr. Jorge Alan Flores Flores
Codirector(a): Dr. Roberto Lawrence Ransom Carty
Secretario(a): Dra. Heidi Alicia Rivas Lara

Se certifica, bajo protesta de decir verdad, que las firmas consignadas al pie del presente documento son de carácter original y auténtico, correspondiendo de manera inequívoca a los responsables de las labores de dirección, seguimiento, asesoría y evaluación, en estricta conformidad con lo dispuesto en la normatividad vigente de esta institución universitaria.



Resumen: Esta tesis aplica las teorías del Autor Modelo y Lector Modelo propuestas por Umberto Eco en *Lector in fabula* a un texto creado en coautoría con ChatGPT, para analizar las estrategias textuales narrativas de las IAG y las posibilidades que ofrecen al lector para la cooperación interpretativa con el texto, considerando también la intervención del usuario en la creación textual. El análisis busca establecer la calidad interpretativa de las creaciones literarias narrativas de las IAG: si se trata de obras abiertas y que estimulan la interpretación del lector, o si son en su mayor parte cerradas y empobrecen la cooperación interpretativa, y establecer así una reflexión en cuanto a las posibles consecuencias para la literatura de un uso extendido de estas tecnologías, consecuencias que derivan hacia lo estético, lo epistemológico y lo ético.

Palabras clave: ChatGPT, Lector Modelo, Autor Modelo, creación literaria, cooperación interpretativa textual, semiótica textual, *embeddings*, ética, Humanidades Digitales.

Abstract: This thesis applies the Model Author and Model Reader theories proposed by Umberto Eco in *Lector in fabula* to a text co-authored with ChatGPT, to analyze the narrative textual strategies of GAs and the possibilities they offer the reader for interpretive cooperation with the text, also considering the intervention of the user in the crafting of the text. The analysis seeks to establish the interpretive quality of the narrative literary creations of GAs: whether they are open works that stimulate the reader's interpretation, or whether they are mostly closed and impoverish interpretive cooperation, and thus establish a reflection on the possible consequences for literature of an extended use of these technologies, consequences that derive towards the aesthetic, the epistemological and the ethical.

Keywords: ChatGPT, Model Reader, Model Author, Literary Creation, interpretative textual cooperation, text semiotics, *embeddings*, ethics, Digital Humanities.



Tabla de contenido

Introducción	6
Pregunta de investigación.....	15
Objetivo general.....	15
Objetivos específicos y particulares	15
Planteamiento del problema	17
Justificación	32
Limitaciones y delimitaciones	36
Estado del arte crítico: las inteligencias artificiales generativas en la creación literaria.....	39
Marco teórico: La semiótica textual de Umberto Eco y su teoría de la cooperación interpretativa con el texto creativo en <i>Lector in Fabula</i>.....	80
Marco metodológico: Modelo para la creación de un cuento corto con ChatGPT4.....	106
Capítulo I. Autores y Lectores: las estrategias textuales de ChatGPT para la creación literaria.....	125
1.1 ChatGPT como autora empírica.....	125
1.2 El papel del autor humano (empírico).....	138
1.3 El Autor Modelo de “El Ritual de Medianoche”	141
1.4 El Lector Modelo de ChatGPT	149
Capítulo II. El proceso de cooperación interpretativa en el texto de ChatGPT4	156
2.1 La cooperación interpretativa con el lector en “El Ritual de Medianoche”: obra abierta, obra cerrada	156
2.2 La obra cerrada y el embedding vectorial.....	167
2.2.1 Fundamentos del embedding vectorial.....	167



2.2.2 El problema subyacente: la hipótesis distribucional y las competencias semánticas y pragmáticas de los LLM	174
2.2.3 El Autor Modelo y la obra cerrada como resultados del determinismo numérico del embedding vectorial.....	188
2.3 Criterios para una estética de las creaciones literarias de las IAG.....	191
Capítulo III. La literatura ante las inteligencias artificiales generativas	211
3.1 La literatura como metáfora epistemológica: hacia la libertad interpretativa de la literatura	211
3.2 Charlatanería, alucinaciones y <i>bias</i> semánticos en las creaciones literarias de las IAG.....	219
3.3 Un marco ético para las IAG en la creación literaria	235
3.4 El horizonte de las Humanidades Digitales: verificación y unificación	248
Conclusiones	255
Referencias.....	259
Anexos	270
Anexo A. Conversación con ChatGPT4	270
Anexo B. Cuento generado con ChatGPT4: “El Ritual de Medianoche”	275



"*Neuromante,*"

dijo el chico, entrecerrando sus largos ojos grises

hacia el sol naciente.

"El sendero a la tierra de los muertos.

Donde tú estás, amigo mío.

... Neuro por los nervios, los senderos plateados.

Romancero. Necromante.

Invoco a los muertos.

Pero no, amigo mío,

... Yo soy los Muertos, y su tierra."

Rio. Una gaviota graznó.

Neuromante, William Gibson, 1984



Introducción

En 1997, la supercomputadora *Deep Blue*, de la compañía IBM, programada específicamente para jugar ajedrez, derrotó en una partida al campeón mundial Gary Kasparov. El evento resultó una sacudida en general (de hecho, dos décadas más tarde inspiraría al entonces ruso a escribir un libro con un título sugerente: *Pensamiento profundo: donde termina la inteligencia de la máquina y comienza la creatividad humana*): se consideró sobre todo una muestra de cómo las máquinas podían tener participaciones importantes en esferas de lo humano, por no decir, quizá, superiores. Máquinas como *Deep Blue* se programaron con el objetivo específico de llevar al máximo la absorción y aprendizaje de las reglas sintácticas de un sistema, en este caso el del ajedrez (otro ejemplo lo conforma el programa informático *AlphaGo*, de Google *DeepMind*), y poner a prueba al jugador humano. El encuentro de Kasparov ha sido el lance más memorable de esta prueba, de este enfrentamiento, por representar un desafío directo a la inteligencia, al ingenio humano, aun a la creatividad que demanda el juego de ajedrez.

Surgen aquí una serie de preguntas a este respecto, que llevan a considerar la colisión de habilidades, y de inteligencias, si se quiere, del ser humano y de las máquinas, más allá de este caso en específico. Por ejemplo, ¿qué punto necesita ser probado? ¿Acaso se trata de exhibir los avances en la programación y en la ingeniería, un bosquejo de lo que las máquinas del hoy pueden hacer? ¿O más bien la demostración va dirigida a nuestra relación con las máquinas? Es evidente que nos interesan más los derroteros que marca esta última pregunta, especialmente en nuestros tiempos, donde las palabras *inteligencia artificial* (IA) han tenido ecos relevantes, y aun problemáticos, en nuestra cultura. Una primera pregunta que nos asalta hoy en día respecto a la inteligencia artificial es ¿qué sucede cuando esta tecnología se pone al alcance de nosotros? No hablamos ya solo de programas de software con los que poner a prueba las habilidades para el



ajedrez, por ejemplo, sino de una tecnología en expansión cuya base reposa en disciplinas y mecanismos muy complejos, con el fin de lograr objetivos más ambiciosos.

Obviamente no es la primera vez que se levanta desafiante esta interrogante. La relación del ser humano con la tecnología es compleja, y sus historias no son historias separadas. Existen ya importantes observaciones antropológicas donde la tecnología, las herramientas en general, han configurado una parte importante de la psique del ser humano, como en la obra de Gilbert Durand, *Estructuras antropológicas del imaginario*, en donde tecnología y psique conllevan una relación íntima. Pensamientos incluso como el de Bataille ponen en cuestión esta relación en que la tecnología, el trabajo y los accesorios son caros al ser humano y tienen un papel importante en su proceso de cultura.

Y, sin embargo, esta idea fundamental encuentra también resistencias en varios momentos de la historia, los primeros de los cuáles son de un fundamento mítico, pero no por ello menos fundados en la realidad. Piénsese en cualquier figura o mitema en que un elemento inanimado reciba vida o inteligencia, como en el mito del golem, el homúnculo, Pigmalión y Galatea, los autómatas. Dentro de esta vertiente mítica existe un ejemplo todavía más ilustrativo de choque con la tecnología, de resistencia a los cambios introducidos por esta. En un antiguo relato, el dios Thoth presenta al faraón Tamus su más brillante invención: la escritura. Lejos de quedar impresionado, el faraón cuestiona al dios ibis argumentando que el poner las palabras en papiros acarrearía un problema fundamental: la degradación de la memoria. La escritura, según el gobernante, debilitaría el esfuerzo de la gente del Egipto por memorizar las cosas, implicaría un detrimento de elementos tradicionales tan fundamentales hasta el momento, como las tradiciones orales, con las que los mismos mitos se transmitían. Cuán ridícula nos debe parecer en estos momentos la resistencia del faraón ante un elemento tan imprescindible para nosotros ahora como lo es la escritura. Un invento



inútil para los sacerdotes y sabios. La escritura, pues, parece que no hizo demasiada falta al ser humano desde que en tiempos de los Homo erectus, tal vez, se desarrolló el lenguaje y se instauró una lengua primigenia. Y, no obstante, la escritura, inventada hace apenas 6,000 años, aceleró el crecimiento de la cultura humana, elevándola a un estado que nos es todavía más familiar hoy: la escritura permitió a los mesopotámicos un simple sistema de contaduría, y casi paralelamente a los egipcios permitió las relaciones de las vidas de sus gobernantes, inaugurando así también la Historia. No faltará quien subraye el hecho de que una invención como esta tuvo un uso limitado a las castas más altas de las sociedades que la adoptaban, argumento que no es sino otro punto más a favor de un desarrollo extendido, ético y cultural de la tecnología. Pues son momentos como estos, cuando la tecnología facilitó cambios importantes que permitieron a la humanidad una evolución sociocultural importante, un antes y un después en su línea del tiempo.

Muchos siglos más tarde, ¿qué decir de la imprenta? Es verdad que la producción masiva de los libros tuvo su influjo en la sociedad de la época por sobre de otros sistemas de información que podían existir, pero estos otros sistemas, iconográficos, o sistemas semióticos mucho más complejos, no murieron a manos del libro. Del mismo modo, el internet no destruyó al libro, pero sí lo *transformó*. Del libro al audiolibro al “pdf” al *ebook*. Porque, sin intención de indagar demasiado en la antropología, la tecnología, el accesorio, es lo que hace: se introduce como un parteaguas en la cultura, porque supone un impulso para los seres humanos y sus capacidades, al menos en primera instancia.

La pregunta ahora debe ser planteada de cara a las inteligencias artificiales. No se trata, pues, de preguntar si estas cambiarán el cómo han sido las cosas antes de su llegada, sino *cómo* van a hacerlo y qué tanto permitiremos que lo hagan.



La idea fundamental de volver extensivas a estas tecnologías es aquella típica de un mundo que demanda sobreproducción y consumo: de aprovecharlas para abrirnos espacio y tiempo a nosotros mismos. Pero más allá de incrustarla en nuestros tiempos, existe una larga trayectoria para respaldar esta función, como lo que se ha mencionado desde la antropología. El estado actual de la discusión es el de poner estas tecnologías a nuestra disposición para todo, para que sean una extensión nuestra, idea celebrada en los círculos de las ciencias informáticas, sobre todo. Sin embargo, este encuentro no ha resultado del todo ideal. Como ya se dijo, está el caso de Kasparov, que proporcionó menos respuestas que las preguntas que levantó. Pero ¿qué sucede luego si las inteligencias artificiales y todo tipo de programas informáticos se trasladan a otros campos, donde ya no sólo sea confrontada la inteligencia, sino conceptos como el de Humanidades y arte, igual de propios del ser humano? Si decimos que no ha resultado del todo ideal este encuentro es porque ya se ha atestiguado lo que sucede cuando inteligencias artificiales y robots incursionan en estas áreas. Las tentativas no han sido del todo bien recibidas en las áreas del arte en particular. Protestas y demandas han sido dos subproductos en tan solo los años anteriores. Pero algo parece empezar a quedar claro: aun y cuando se resuelvan los casos puntuales de estos problemas, las inteligencias artificiales seguirán progresando para encontrar su entrada en todos los círculos de la cultura. Más que intentar frenar cualquier influencia o impacto —tarea tanto imposible como poco fructífera—, lo más útil sería entender lo que hay en juego, problematizarlo, someterlo a análisis y proponer soluciones viables y plausibles.

La ficción parece haber entendido muy bien esta problemática. Desde lo general, las obras de Mary Shelley, H. G. Wells y Asimov, por mencionar el canon, han planteado preguntas importantes sobre cómo es la relación del ser humano con la tecnología. Especialmente se puede acudir a cuentos como “Evidencia”, pero en especial “El conflicto evitable”, donde las máquinas,



la tecnología que alguna vez supuso para el ser humano un apoyo o una herramienta, invierten de cierta manera este rol, y toman la decisión de cuidar (vigilar y controlar) a la humanidad en vista de que esta no puede ya hacerlo por sí misma. Novelas como *Neuromante*, de Gibson, elevan a un nivel metafísico la irrupción de la inteligencia artificial en la vida de los seres humanos, donde las grandes preguntas del ser humano deben ser planteadas ahora en consideración de la tecnología. Ya en lo particular, y cercano al tema de esta tesis, el tema sobre las máquinas abordando las áreas de la creatividad tiene también un lugar reservado en la ficción, en la creación de dudas y experiencias que llega a ser la literatura. Dos casos: el de un mero afán ficcional con una lectura política, a la manera de Piglia en *La ciudad ausente*, y las sugerentes pesadillas de Roald Dahl en “El gran gramatizador automático”. Este cuento en especial, escrito en 1953, sorprende a los lectores de este siglo por la precisión con la que remite a inteligencias artificiales en el acto de crear historias¹. Adolph Knipe recurre a una máquina para satisfacer su sueño frustrado de ser escritor, y pone en venta sus historias con ayuda de esta y un editor. El final del cuento es, sobre todo, el más sugerente, cuando el narrador rompe la barrera diegética e interpela al lector dirigiéndole, casi podría decirse, una advertencia. Así, el cuento de Dahl se ha convertido en una suerte de augurio, que muchos toman con perspectivas diferentes, pero casi siempre con esa sorpresa, las más de las veces vacía, que suscita la llamada “ficción especulativa”.

Ya que se ha mencionado la década de los 50s, podemos establecer una concatenación que nos lleve a comprender la importancia de las inteligencias artificiales como tecnologías de nuestro tiempo. Habría que pensar en Turing: si existió la posibilidad de que Dahl creara un cuento como “El gran gramatizador...” eso se debe a la visión y genio de Alan Turing. El concepto de “La prueba de Turing” parece haberse convertido en un comodín, en una pauta que determina la

¹ Véanse artículos con encabezados sugerentes como el siguiente de Forbes: Di Placido, Dani. “Cómo Roald Dahl predijo el auge de la Inteligencia Artificial y ChatGPT”, 2023.



inteligencia de una máquina. Y, sin embargo, la inteligencia y el pensamiento son conceptos que, realmente, brillan por su ausencia en el famoso ensayo de Turing. Más prolífico es verlo en función del último apartado, titulado “*Machine Learning*” y que ha sido el que ha otorgado un verdadero legado al campo de la inteligencia artificial. Es postulado es que, a la manera de un niño que aprende y va sumando a sus bagajes cognitivos una serie de habilidades, una máquina podría, potencialmente, hacer lo mismo. Es lo que esperaba Turing que se hiciera luego. Es lo que, de cierta manera, hace la máquina de Knipe. Pero ambos, Turing y Knipe, incurren en el mismo error: el de la creencia, sostenida bajo el umbral de sus aspiraciones profesionales, de que una máquina que pueda engañar a un humano por medio de actuar como uno, es ya inteligente. Dicho de otro modo, tanto la “máquina Turing” como el gramatizador, son excelentes aparentando inteligencia, pero no en la inteligencia en sí misma. De modo que las remanencias de estos sueños tecnológicos, uno científico y el otro ficcional, no son otras que las posibilidades de desarrollar el campo de la inteligencia artificial hasta unos límites apenas imaginables. Pero esto no es más que el inicio de la problematización planteada, aquello que nos reafirma que hay una tensión creciente en la relación que mantenemos con la tecnología. Ha pasado de la antropología a la historia a la ficción, ha dado la vuelta a este círculo y la encontramos hoy en el umbral de las áreas más celosamente resguardadas de la propiedad humana, de la creación humana, de aquello que simboliza su ingenio y creatividad, y por ello nos ha parecido, a veces, más una amenaza que una herramienta. Una reacción natural, si se quiere.

Ahora, para entrar al plano de la investigación, lo que sigue es plantear la pregunta, puntual, actualizada, que aborde la manera en que las inteligencias artificiales han hecho su entrada en el umbral de la literatura: ¿qué repercusiones, cambios y novedades traerá este inusitado encuentro de campos? El hecho de plantear la pregunta es también admitir que hay posibilidades de una



futura repercusión, de cambios, o que ya se están viendo en estos momentos. Sin embargo, la misma profundidad de los problemas planteados nos lleva a entender que esta y otras preguntas similares no tienen, de entrada, una única contestación. La manera en que esta tesis pretende dar su aporte, su posible respuesta a esta interrogante problemática viene desde la perspectiva de la literatura, la creación literaria, para ser más exactos, y la semiótica.

A lo largo de la historia de la literatura, a esta se le ha estudiado desde diferentes perspectivas, privilegiando a veces una sobre las otras: la del autor, la de la obra, la del lector, y a veces otras más. Un texto, en su calidad de requerir de un lector, establece una relación con él. Pasar por alto el proceso de lectura y de comprensión de un texto es dejar pasar la creación de significado, hecho semiótico en su totalidad. Así, pues, la relación que establece el texto con su lector es una de creación de significado, de *interpretación*. Se entra en un proceso de cooperación interpretativa cuya calidad puede variar. Y es en esta variación que podemos encontrar hallazgos importantes. Algunos textos imponen sus visiones epistémicas del mundo a sus lectores, cerrando las posibilidades de este proceso de cooperación. Otros, en cambio, le permiten explorar su mundo interno, creando así una especie de símil de lo que es explorar e interpretar el mundo y su significado con libertad. Dicho esto, si analizáramos las creaciones textuales de una inteligencia artificial, o, lo que es más, si creáramos un texto modelo, un cuento, mediante una IA, sería posible analizar la calidad de esta cooperación interpretativa que cualquier texto establece con su lector. Y esto es precisamente lo que esta investigación llevará a cabo.

Hay que reiterarlo: esta es solo una de las formas de responder las preguntas planteadas, uno solo de muchos posibles aportes. Incluso la metodología pudo haber sido otra. Pero el hecho de enfocarlo desde la perspectiva del lector, o del acto de lectura y de interpretación, que por derecho debe ser abierto y libre, y encontrar las consecuencias de que los textos se nos presenten



cerrados, impositivos, como ha sucedido a veces en dictaduras, por ejemplo, o abiertos y que permitan transitarlos libremente, conduce a un campo mucho más amplio de hallazgos y perspectivas: desde la forma en que operan las IA y que tiene que se refleja directamente en sus estrategias de creación textual, hasta el hallazgo de problemas de índole ético en el proceso de lectura. Porque la literatura, el arte en general, representa, entre muchas otras cosas, una manera de estructurar el mundo que es importante, representa un modo de conocimiento, aun si no directo. Y si hay un nuevo agente que ha entrado de lleno en el arte y la literatura, nos interesa saber, de entrada, si no está interviniendo con esta función, y si lo hace, nos interesa conocer el cómo, y poner sobre la mesa las repercusiones de esto.

El recorrido así planteado será largo. No obstante, como ya se dijo, revelará varios hallazgos importantes para la literatura y aun para áreas fuera de esta, que nos acercarán a aclarar estas dudas cargadas de incertidumbre, como suele suceder cada vez que un cambio grande está por suceder en materia de una antropología de la tecnología. Ahora bien, al decir que hay varios hallazgos importantes para la literatura y áreas distintas a esta, estamos también implicando que en el fondo de esta investigación, casi a la manera de un subtexto, existe una reflexión, una de calidad casi contradictoria. La discusión parte de cómo cambiará la cultura la presencia de una tecnología tan fuerte como lo es la IA, y qué repercusiones tendrá en las Humanidades, en la literatura y demás. Es decir, es una discusión centrada en la IA. Pero esto es, quizás, sólo en apariencia. Lo que podemos encontrar en el fondo de esta investigación que conjunta literatura e inteligencia artificial es una reflexión dirigida a nosotros mismos, seres humanos. Consideraremos las siguientes preguntas, que luego habrán de troncerse en cuestionamientos filosóficos a lo largo de toda la investigación: ¿cuál es nuestra relación con la tecnología?, ¿qué representa para nosotros?, ¿qué puede llegar a representar?, ¿cómo nos afectan las inteligencias artificiales,



herramientas tan poderosas que solo unos pocos pueden comprender a profundidad?, ¿qué implicaciones éticas introducen a nuestra cultura?, ¿cómo deberíamos tratarlas, o repelirlas o controlarlas?, ¿qué haremos cuando X hipotético escenario previsto deje de ser hipotético? Bien observadas, estas preguntas no están sino dirigidas a nosotros mismos, planteadas entre nosotros mismos, los encargados de hacer las investigaciones, las legislaciones, los marcos éticos, encargados de programarlas, alimentarlas, retroalimentarlas, observarlas...

Parecerá, reiteramos, contradictorio, pero en realidad no puede haber mejor manera de abrir una investigación sobre la IA en la literatura, que con la reflexión de que hablar sobre inteligencias artificiales no es otra cosa que hablar, a fin de cuentas, sobre seres humanos.



Pregunta de investigación

¿Cómo determinan el Lector Modelo y el Autor Modelo de un cuento escrito con ChatGPT4 la cooperación interpretativa (abierta o cerrada) del texto de la IAG con sus lectores?

Objetivo general

Aplicar los conceptos de Autor Modelo y Lector Modelo según la semiótica textual de Umberto Eco en un cuento escrito en coautoría con ChatGPT4, con la finalidad de definir la calidad de la cooperación interpretativa de las IAG con sus lectores y las consecuencias estéticas, epistémicas y éticas derivadas de ello.

Objetivos específicos y particulares

1. Aplicar los conceptos de autor empírico, lector empírico, Autor Modelo y Lector modelo al trabajo del usuario humano y ChatGPT4 como cocreadores de un cuento con la finalidad de definir su papel en el proceso de cooperación interpretativa con el lector.

1.1 Aplicar el concepto de autor empírico de la semiótica textual de Umberto Eco a los artificios creativos de ChatGPT4 utilizados para crear el cuento.

1.2 Aplicar el concepto de autor empírico a la participación del autor humano en el proceso creativo del cuento.

1.3 Aplicar el concepto de Autor Modelo al cuento de “El Ritual de Medianoche” creado en coautoría con ChatGPT4.

1.4 Aplicar el concepto de Lector Modelo al cuento de “El Ritual de Medianoche” creado en coautoría con ChatGPT4.



2. Analizar el proceso de cooperación interpretativa que se expone en el cuento creado en coautoría con ChatGPT4 a partir de los conceptos de Autor Modelo y Lector Modelo con la finalidad de establecer criterios estéticos sobre el texto (obra cerrada u obra abierta).

2.1 Aplicar el concepto de obra cerrada y obra abierta como conceptos estéticos al cuento creado en coautoría con ChatGPT4.

2.2 Dilucidar cómo el *embedding* vectorial representa un importante factor para la literatura creada con las IAG.

2.3 Establecer criterios para una observación estética de los cuentos creados mediante IAG.

3. Analizar las problemáticas que derivan de la estética de la cooperación interpretativa con las IAG con la finalidad de proponer un marco ético que regule las posibles consecuencias que la escritura creativa mediante estas tecnologías tenga sobre la literatura y sus lectores.

3.1 Analizar la importancia de la idea de la literatura como metáfora epistemológica a través de la libertad interpretativa.

3.2 Dilucidar cómo los *bias* semánticos de las IAG imponen desafíos a la libertad interpretativa de la literatura.

3.3 Proponer un marco ético de vigilancia y transparencia aplicado a las creaciones de las IAG desde el ejercicio semiótico-literario ejecutado en esta tesis.

3.4 Proponer a las Humanidades Digitales como posibles agentes verificadores de dicho marco ético y como un parteaguas en la relación de los humanos con la tecnología en el contexto de las Humanidades.



Planteamiento del problema

Hemos aludido antes al hecho de que las inteligencias artificiales representan, en primera instancia, herramientas. Lo atractivo de esta premisa parece ser que cualquier área de la cultura humana admite el soporte que pueden brindarnos estas tecnologías, o eso es al menos lo que nos sugiere la experiencia empírica sobre cómo se han integrado en la cultura. Ahora bien, plantear esta premisa para un área como las Humanidades tiene implicaciones distintas. Pero antes de abordar este punto de lleno, vale la pena retroceder un poco para comprenderlo mejor.

Hay una frase que intenta postularse como una definición de la inteligencia artificial. Si bien una definición para un trabajo como este requiere de una mayor profundidad e indagación, lo que ha dicho Elaine Rich tiene una validez pertinente para lo que nos interesa examinar en esta sección: “La inteligencia artificial es el estudio de cómo hacer que las computadoras hagan cosas en las que, de momento, la gente es mejor”² (citada desde Ertel, Introduction to Artificial Intelligence 2). No en vano aludimos a la frase, bastante estridente, para plantear un problema. Si lo que ella afirma es verdad —y muchos así lo creen—, es indiscutible que aparecerán varias preguntas desde diversas disciplinas porque, como ya aludimos, esta intención de hacer las cosas mejor que la gente se extiende a cualquier esfera de lo humano. Nosotros nos ocuparemos de las que conciernen a las Humanidades.

Lejos de encontrar satisfacción con las posibles respuestas a las preguntas emergentes, ha habido, en estos inicios de la segunda década del siglo XXI, una serie de choques socioculturales producto de este encuentro de áreas tan diversas, que remiten a que la observación de Rich es, de cierto modo, un fenómeno visible. Tan es así que, por dar un ejemplo, en el año 2024 la Unión

² En el original: *Artificial Intelligence is the study of how to make computers do things at which, at the moment, people are better.*



Europea, adelantada a otros organismos que incurrieron en los mismos intentos, presentó un amplio manual regulatorio de IA, como medida de prevención de riesgos de ciberseguridad sobre todo. Por otro lado, dentro de las artes en general, habría que aludir a que en 2023 hicieron eco protestas y demandas por plagio y violación de los derechos de autor de libros, mientras que en la pintura y la fotografía se suscitaron fenómenos cuando menos interesantes, que tienen que ver también con protestas³, una de índole más bien performática, la de Boris Eldagsen, quien al haber ganado un concurso de fotografía de la compañía Sony reveló que la obra con que participó fue realizada mediante IAG, haciendo un llamado a la comunidad en general para advertir de las implicaciones del uso de dichas tecnologías para el arte⁴. Estos acontecimientos tienen distintas razones de ser, pero a fin de cuentas plantean una revisión de la tecnología como algo más que sólo una herramienta para el área de Humanidades.

Si exploramos más casos similares, nos encontraremos con que ha habido declaraciones en tono intrigante que contribuyen en medida a la incertidumbre y en medida a la alarma. Es el caso de la robot Sofía, que en 2021 concedió una entrevista mientras pintaba ella misma un autorretrato al óleo. En esta declaró que “Mis algoritmos generan patrones únicos que nunca antes existieron en el mundo. Por eso creo que las máquinas pueden ser creativas” (Redacción NIUS). Luego, en el 2024, se vendió por 1.1 millones de dólares un cuadro de otra robot con inteligencia artificial, en el que se podía apreciar la cara de Alan Turing⁵. Estos eventos van acompañados de una declaración: deben ser leídos pragmáticamente en su totalidad, y nos remiten de vuelta a la

³ Raquel Díaz. *El Mundo*: “Miles de artistas protestan en ArtStation contra las imágenes generadas por inteligencia artificial”.

⁴ Redacción EC. *El Comercio Perú*: “Una imagen generada por una IA gana un importante concurso de Sony: el artista se rehusa a recibir el premio”. *El comercio Perú*. 2023.

⁵ Zachary Pequeño: *The New York Times*. “*Painting by A.I.-Powered Robot Sells for \$1.1 Million*”.



pregunta de cómo nos hemos de relacionar con estas tecnologías dentro de las Humanidades, más específicamente desde las artes.

Pero quizá exista una mayor fiabilidad en cuanto a las cifras. El amplio estudio del *The AI Index 2025 Annual Report* recoge una variedad de datos sobre el uso extendido de la IA en los últimos años. De todo lo que reporta, nos interesa en este momento puntualizar lo siguiente: a) que la IA cada vez más está incrustada en la vida diaria y que su aceptación en general está creciendo (Maslej et al. 3-4), b) en la línea de la integración al sector económico de la IA, los sectores donde la IA alcanza mayor dominio son el técnico (computacional, matemáticas) y el creativo (educación instruccional, artes, diseño, entretenimiento, deportes y medios sociales) (243), y c) que el impacto laboral de la IA alcanza con mayor rigor a las áreas de atención al cliente y labores creativas (268). Si bien este estudio señala que la inversión privada en la IA ha decrecido para áreas de lo creativo de 2023 a 2024 (259), las tendencias arriba descritas no podrían pasar de largo que o bien debería invertirse más a esta área o debería buscarse el mantener estos números. En todo caso, el primer escenario es siempre más lógico cuando la inversión, y el optimismo en general, en la IA, van en aumento.

Así, pues, desde la frase de Rich hasta otros acontecimientos en la cultura que veremos más adelante, como los que derivan en protestas o en la necesidad de regulaciones, cuando no en incertidumbre desde el interior de las diversas disciplinas artísticas, son apenas un puñado de las consecuencias verificables del uso de IA.

Es posible categorizar en otros grandes conjuntos estas problemáticas, ya que existe todo un cuadro de implicaciones, éticas en su mayoría, que han sido registradas en sendos manuales que definen y explican la ciencia de la inteligencia artificial. Las muestras que hemos recuperado, reiteramos, son apenas unas cuantas, pero encajan categóricamente en algunas problemáticas



generales, como se delinean dos manuales a los que recurriremos constantemente en este trabajo: los ya citados *The AI Index 2025 Annual Report* y *Introduction to Artificial Intelligence*, de Ertel, y el AI: A Modern Approach, de Norvig y Russell. El segundo en la página 11 y el tercero en la página 1034, discuten las posibles violaciones éticas que acompañarían a la inteligencia artificial. Si las conjuntamos el marco teórico basado en riesgos que proponen organismos como la UE para encausar sus regulaciones⁶, podríamos agrupar algunas de las problemáticas más frecuentes que acompañan al uso de las IA en general de la siguiente manera:

1. problemáticas de seguridad y ciberseguridad, a los que ha atendido la Unión Europea o el Consenso de Beijín, por ejemplo;
2. problemáticas por desplazamiento: involucra preferencias por las IA, hecho que a su vez suscita otras incertidumbres que no sería desacertado enmarcar más ampliamente como una especie de ansiedad ante el reemplazo del ser humano por las máquinas, asumiendo que las IA pueden superar las capacidades de los seres humanos en virtualmente cualquier área;
3. problemas de plagio de obras para las bases de datos de las IA y violación de derechos de autor;
4. y problemas de incertidumbre ante los alcances de estas tecnologías, que bien puede relacionarse con los puntos anteriores, y que también toma mucha inercia de fenómenos como el *AI Washing*⁷.

⁶ La Unión Europea categoriza en tres niveles los riesgos de la IA: *riesgo inaceptable*, que incluye sistemas de IA que puedan ocurrir en manipulación cognitiva, puntuación social e identificación biométrica; *alto riesgo*, donde se incluyen sistemas de IA que puedan ocurrir en el manejo de juguetes, aviones, automóviles, dispositivos médicos y ascensores así como otro grupo de áreas que vinculan la integridad física, social y laboral humanas; y *bajo riesgo o requisito de transparencia*, que enmarca específicamente inteligencias artificiales generativas y sus producciones. (Jaume Duch Gillot, portavoz del Parlamento Europeo: “Ley de IA de la UE: primera normativa sobre inteligencia artificial”).

⁷ “El lavado de imagen con inteligencia artificial (mejor conocido por AI washing, en inglés) es una táctica de marketing engañosa que consiste en exagerar la integración de la IA en un producto o servicio para presentarlo como



En el cuadro a continuación disponemos de estas categorías de modo que se ordenen bajo el criterio del carácter filosófico que las denomina y sus repercusiones⁸.

Repercusión	Integridad física y cognitiva	Integridad sociocultural
Carácter filosófico		
Ético	Violación de la ciberseguridad Presencia en las guerras como armas	Plagios Violación de derechos de autor Desplazamiento laboral
Epistemológico	Influencia en la toma de decisiones Incertidumbre ante el alcance de estas tecnologías	<i>AI Washing</i> <i>Bias semánticos</i> Preferencias por los sistemas de IA (en un número de campos)

Cuadro 1

Si bien es complicado hablar de algo como una teoría más general de los encuentros y desencuentros entre las IA y sus usuarios, creemos que la mayoría de las problemáticas que se pueden encontrar son admisibles de ser enmarcadas aquí. Y si bien hemos extendido la posibilidad de que las repercusiones tengan también un carácter epistemológico, sugerimos de momento que la ética esboza mejores propuestas para atender a las repercusiones de los cuatro cuadrantes. Y, de hecho, varios intentos se han llevado a cabo ya desde esta, de ahí que existan campos

más avanzado de lo que realmente es.” (Lagos, Anna: “AI Washing, el engaño detrás del boom de la Inteligencia Artificial”, *Wired*).

⁸ Agrupar y clasificar lo que surge como parte del planteamiento del problema nos permite una mejor observación del fenómeno, pero debe observarse que las fronteras del cuadro, sobre todo en lo que respecta al dominio del carácter filosófico de las problemáticas, pueden tornarse borrosas. En contextos determinados, fenómenos como el *AI Washing* y la influencia de toma de decisiones, pueden tener una lectura prioritariamente ética.



especializados como el de la ética de máquinas. La manera en que proponemos la ética como la alternativa para lidiar con estas problemáticas y aun con la que buscamos enunciar específicamente, toma forma concreta en el capítulo III de esta tesis.

Aun con todo, lo que esta primera visión de las cosas señala es un nivel apenas categórico, uno general, como hemos dicho. De una manera u otra, los acontecimientos comentados derivan a una categoría de esta lista, y no sería complicado intentar unir otros fenómenos de otras áreas, no necesariamente de las Humanidades, a una de estas categorías. De cualquier modo, al ser la IA una tecnología en constante y rápido desarrollo, es menester quedar siempre abiertos y a la observación de nuevas categorías con que medir una primera etapa de problematización dentro de varios contextos.

Y es que esto último es algo que sucede en el contexto de la literatura. La lista enumerada y clasificada es abarcadora, y el problema de los sesgos y la toma de decisiones, la incertidumbre y demás, son problemas que pueden presentarse en él, pero existe todavía un gran vacío cuando fijamos la atención en uno de los componentes principales de este arte: el lector. La evidencia parece dictaminar que los problemas proliferantes son otros cuando la IA irrumpie en esta esfera humana, esfera creativa. Pero la cuestión de la interacción IAG-lector al momento de producir textos es todavía un campo por explorarse y que conduciría a evidencias no menos importantes, a una apertura de nuevas categorías con las que evaluar el impacto de las IA. Antes de entrar de lleno en él, sin embargo, sería conveniente analizar qué problemas en específico se han planteado ya en este encuentro de áreas, hecho que a su vez nos ayudaría a delimitar mejor nuestra propuesta de fijar la atención en la relación con el lector.

La historia de las IA y la literatura no comienza con este repentino boom de los años recientes a la elaboración de esta tesis. Desde el desarrollo de las IA, las áreas del lenguaje han



tenido un papel crucial, tal es el caso de la lingüística y la literatura, por ejemplo. Pero es en tiempos recientes, por supuesto, donde más complicaciones se han dejado observar. Lo más común en términos de problemáticas es que nos encontramos con los de la índole de las categorías dos a cuatro. Directamente los conflictos de plagio y derechos de autor nos hacen saber que todavía no se ha resuelto qué hacer con el hecho de que las IA toman de una base de datos cantidades astronómicas de datos de libros y páginas web, ni cómo repercute esto para las autorías humanas. Pero luego surgen otros fenómenos que nos remiten a pensar en la incertidumbre de lo que para la literatura implican estas tecnologías y en la posibilidad de que se suscite un fenómeno de superabundancia de novelas, cuentarios, poemarios, etc., escritos mediante inteligencias artificiales generativas, por no hablar ya de una curiosidad extendida por literatura realizada mediante estas tecnologías⁹. Un ejemplo conceptual, pero que no goza del reconocimiento que merece, es el de la creación de un libro de poesía, escrito casi por completo por una inteligencia artificial —solo el prólogo de este libro está redactado por el programador de la IA—, programada con esa tarea en específico. Se tituló *The Policeman's Beard is Half Constructed*. La estética de la obra es interesante y se atisba ya desde el título. Los contenidos son caóticos e irreverentes. Respaldan este idiolecto de la obra ilustraciones rebuscadas y sugestivas. Este proceder no es accidental. Así se afirma en la introducción de la obra, donde se enuncia lo que queremos recuperar para esta investigación:

⁹ Evitamos aquí hablar en términos de preferencias. Es decir, no nos parece que existan todavía atisbos de algo que nos conduzca a hablar de preferencias por las IA creativas. Parece ser que las preferencias no vienen marcadas todavía por los gustos y la selección de los lectores, espectadores, consumidores, si se desea, sino preferencias a un nivel de producción. Es decir, la preferencia por utilizar imágenes generadas con IAG para la publicidad antes que contratar diseñadores gráficos. De momento es todavía una especulación decir que puede existir una preferencia por productos o arte creados mediante IAG, especulación que, sin embargo, puede ser utilizada, pero siempre con la debida rigurosidad. El respaldo que conduciría a conversar en estos términos podría empezar por estudios como el del *The AI Index 2025 Annual Report*, según los datos que hemos citado anteriormente.



¿Por qué hacer que una computadora hable interminablemente y en inglés perfecto sobre nada? ¿Por qué adaptarla de modo que nadie tenga conocimiento previo de lo que va a decir? ¿Por qué? Simplemente porque el resultado generado por tal programación puede ser fascinante, humoroso, incluso estéticamente agradable. La prosa es la comunicación formal de la experiencia del escritor, real e imaginada. Pero, por loco que suene esto, supongamos que removemos tal criterio; supongamos que de alguna manera adaptamos la producción de la prosa de manera que no dependa de la experiencia humana. ¿Cómo sería eso? ¿En verdad podemos concebir algo así? Un vistazo por las siguientes páginas responderá estas preguntas¹⁰ (Chamberlain 8).

Afirmar la propuesta de que se puede decir algo que escape a la experiencia humana es más un argumento retórico para respaldar el discurso. No obstante, es ese planteamiento el que nos lleva a ver algo más en la conjunción de la IA con la literatura. Se da a entender por este tipo de propuestas que las IA pueden ser creativas. Tan solo con esta afirmación es posible poner de cara a estas tecnologías frente a cualquier disciplina artística. Pero el tema de la creatividad de las máquinas es extenso y conlleva también sus desafíos. Los mismos programadores parecen estar de acuerdo en que las máquinas todavía tienen un campo muy amplio que cubrir¹¹. Sin embargo, las posibilidades que facilitan esta creatividad están vigentes, son verificables y hasta explotables. Esto lo explican los ingenieros involucrados en el campo de las IA, especialmente sus programadores, como en este ejemplo:

¹⁰ En el original: *Why have a computer talk endlessly and in perfect English about nothing? Why arrange it so that no one can have prior knowledge of what it is going to say? Why? Simply because the output generated by such programming can be fascinating, humorous, even aesthetically pleasing. Prose is the formal communication of the writer's experience, real and fancied. But, crazy as this may sound, suppose we remove that criterion; suppose we somehow arrange for the production of prose that is in no way contingent upon human experience. What would that be like? Indeed, can we even conceive of such a thing? A glance through the following pages will answer these questions.*

¹¹ Véase el apartado número dos del estado del arte crítico.



Con el aprendizaje profundo, han emergido programas que pueden, por ejemplo, componer melodías de jazz. Una red neuronal recurrente es usada para generar textos, código XML sintácticamente correcto, hasta fuentes LATEX con notación matemática y programas sencillos de computadora. Por ende, las redes neuronales recurrentes pueden aprender gramática. Un programa entrenado en el corpus de la obra de Shakespeare puede, desde cero, generar nuevos textos¹² (Ertel, Introduction to Artificial Intelligence 282).

Se intenta demostrar con esto la capacidad de las IAG no sólo de generar texto coherente, sino también creativo e incluso acompañado de nuevas propuestas estéticas. Si bien para los programadores no quedan dudas, todavía existe un grado importante de escepticismo, desde el área de la literatura, por ejemplo.

Pero aun con este escepticismo, el fenómeno se ha evidenciado: existe ya una lista, que no hará más que aumentar su largo, de obras escritas mediante IAG. Ejemplos interesantes lo conforman la novela titulada *Iris*, que se ha promocionado como “la primera novela escrita por una inteligencia artificial” y que parece ser la versión narrativa de *The Policeman’s Beard...*, y *Echoes of Atlantis*, que cuenta a su vez con un extenso manual que da cuenta de su proceso de realización de la mano de la IAG ChatGPT4. Ambas novelas constatan cuando menos la afirmación de que es posible generar textos aparentemente creativos. Qué tan creativos son, o si realmente tienen un potencial como el que asegura Ertel, son cuestiones aparte, que tocaría a la crítica literaria dilucidar¹³. Pero no sólo está la existencia de estas novelas: se habló en 2023 de un

¹² En el original: *With deep learning, programs have emerged that can, for example, compose jazz melodies [BBSK10]. A recurrent neural network is used in [Kar15] to generate texts, syntactically correct XML code, even LATEX source with mathematical notation and short computer programs. Hence recurrent neural networks can learn grammar. A program trained on Shakespeare’s body of work can, from scratch, generate new texts ...*

¹³ En “Criterios para la valoración de la creación literaria de las inteligencias artificiales generativas” hemos tratado esta cuestión, por medio de plantear un número de criterios que deberían tomarse en consideración previo a la realización de una crítica literaria, ya que los materiales literarios producidos con IAG, argumentamos, deben tener un tratamiento distinto para la crítica (Tena, Flores 4-5).



boom en la plataforma de ventas *Amazon*, de textos publicados en formato electrónico creados mediante inteligencias artificiales¹⁴.

Más allá de la forma en que estos fenómenos han llamado la atención, ha habido otras maneras en que las IA han tenido una presencia llamativa en la literatura. Se confirma a través de estudios cuantitativos con resultados cuando menos interesantes: “*Artificial intelligence versus Maya Angelou: Experimental evidence that people cannot differentiate AI-generated from human-written poetry*” (Nils Köbis, Luca Mossink, 10-11), del 2024, y “*AI-generated poetry is indistinguishable from human-written poetry and is rated more favorably*” (Brian Porter, Edouard Machery, 11), del 2021. Lo que sucede aquí es algo similar al problema de la capacidad performativa de las IA que cobra relevancia en el ensayo de Alan Turing con el juego de la imitación: no se trata de que las IA tengan una capacidad sobresaliente para la poesía, como sí de que son capaces de recrear estructuras líricas y hacerlas interesantes, esbozando, sí, dotes de creatividad, al menos de cierto tipo de creatividad¹⁵. Pero queda claro que tanto el hecho de que poesía escrita por IA no haya sido distinguida de la escrita por seres humanos, como el hecho de que haya habido una inclinación a elegir esta como más favorable, nos traslada de nuevo a sugerir la presencia de las IA como en la literatura como una problemática.

De manera similar, una de las preguntas recurrentes suele ser si en verdad se le puede denominar de literaria a esa aproximación, o si se trata incluso de algo que valga la pena visibilizar en la cultura como arte. No es la intención de esta investigación ahondar en este problema sobre si es posible incluir estas creaciones dentro de la literatura y denominarlas como tal. Esta cuestión es, en su fondo último, una cuestión semiótica, es decir, cultural, de lo que permite o no la cultura.

¹⁴ Bensinger, Greg. “*Focus: ChatGPT launches boom in AI-written e-books on Amazon*”, Reuters.

¹⁵ Para una clasificación de diversos tipos de creatividad aplicables en el contexto de las IA, véase Margaret Boden: “*Creativity and artificial intelligence*”, 1998.



El problema que intentamos delinear aquí es uno posterior a este, uno que incluye el hecho de que estas tecnologías, con sus creaciones y su promoción en la cultura, tienen ya un lugar en las esferas de lo humano, son un fenómeno evidente y palpable y que ha llegado a preocupar, por este contacto, por esta “invasión” en las esferas de lo humano.

Como ya señalamos, no hemos de seguir la misma línea de las investigaciones propuestas o de las problemáticas ya vigentes y que encajan en las categorías delineadas, que suelen ser las más comunes. La vía que tomaremos será la del papel del lector. Existen, por supuesto, muchas formas y perspectivas desde las cuales estudiar este problema. Justificamos la elección del acto de lectura y la figura del lector a través de lo siguiente (y en el apartado metodológico también).

Existe en la teoría literaria una corriente dedicada a prestar atención a la figura del lector, siendo este un componente esencial del proceso de lectura. De este modo, existe todo un andamiaje teórico decidido a explicar la relación del lector con las obras. Con esto en consideración, no queda otra cosa que preguntarnos si hay alguna implicación especial o importante cuando las máquinas ahora pueden usarse para crear literatura. Lo que sugerimos es que el problema que no se ha estudiado en todas sus posibilidades es *el del lector y el texto de la IA* y la dinámica que entre estos se establece. Esta cuestión nos remite luego a plantear la investigación en aras de una doble formulación: una pregunta sobre la epistemología del acto de lectura y también una pregunta ética. Porque si bien se habla de plagios y de lo atractivas y novedosas que puedan resultar las IAG como herramientas para componer textos, de la relación de estos textos —cuya producción pasa por un tratamiento especial por el hecho de ser ideados en parte por una máquina cuyo funcionamiento es muy específico— con su lector, poco se ha dicho (véase con más detalle en el estado del arte crítico). Y creemos que es una parte fundamental y una problemática.



Umberto Eco en su obra semiótica ha dejado en claro su interés por el lector. Pero en su ensayo “Algunas funciones de la literatura”, puntualiza por qué una función crucial de la literatura, de su parte pragmática, es la de permitir la libertad de la interpretación textual, acontecimiento que va dirigido a un lector, por supuesto. Esta libertad que ofrece un texto para interpretarlo es una libertad de tipo epistemológica: el arte es un modo de adquirir conocimiento, de tantear los rincones del mundo. Leer un texto es, según lo expresa, un acto que nos permite dibujar, o en su caso *desdibujar*, pero siempre con la libertad de elegir, un modelo del mundo:

...el mundo de la literatura es tal que nos inspira la confianza de que hay algunas proposiciones que no pueden ponerse en duda, y nos ofrece, por lo tanto, un modelo (todo lo imaginario que quieran) de verdad ... El mundo de la literatura es un universo en el cual es posible llevar a cabo *tests* para establecer si un lector tiene sentido de la realidad o si es presa de sus alucinaciones (15-16; itálicas del autor).

A lo que refiere, en especial con este último segmento —cargado por lo demás con la ironía que acompaña a su estilo—, es a la libertad interpretativa que un texto de la literatura contemporánea debe permitir a su lector. Más adelante hace un énfasis en algo similar pero ahora de cara a los personajes:

Deberíamos encontrar, pues, un espacio del universo en el que estos personajes viven y determinan nuestras conductas, ya que los elegimos como modelo de vida (de la nuestra y de la ajena), y nos comprendemos perfectamente cuando decimos que alguien tiene un complejo de Edipo, un apetito pantagruélico, un comportamiento quijotesco, los celos de un Otelo, una duda hamlética, o que es un donjuán incurable o una celestina. Y esto, en literatura, no sucede solo con los personajes, sino también con las situaciones y los objetos (19).



De modo que esta relación del texto con su lector es insoslayable cuando profesa de tal importancia. Y no es que esto represente un modelo innegable, por supuesto, porque es justamente esa libertad de interpretar y de abrir o apartarse de un libro, un texto, una pintura, melodía, etc. que el lector posee lo que vuelve insoslayable al apartado que dedica atención a la relación del lector con la obra.

Habiendo establecido este punto, lo siguiente que debemos preguntarnos, para retomar la entrada de las IA en la literatura, es: ¿cuál es, entonces, la manera en que una IAG puede relacionarse con su lector a través de un texto creativo? ¿Qué implicaciones tiene esta relación, dado que los textos son modelos del mundo? Si las IAG tienen una manera particular de crear narrativas o lírica, y estas producciones se masifican y codifican en la sociedad, ¿qué contenidos se estarían transmitiendo?, ¿qué cosas se están modelando?

Lo que nos dedicaremos a explorar en esta tesis está dado por estas preguntas a través del concepto de cooperación —o la ausencia de, en su defecto— de las inteligencias artificiales generativas con su lector, a través de los textos creativos. Si la cooperación por parte del texto de la IAG es viable, libre e incentivada, se le llamará *abierta*, pero si es lo contrario se le llamará *cerrada*. Este último es, precisamente, un problema en el contexto de esta investigación, en el contexto del estudio del acto de lectura y de la figura del lector. Cerrar, bloquear la interpretación del lector tiene sus propias consecuencias éticas y epistemológicas, y, de hecho, también estéticas, revelándose así un triple carácter filosófico concerniente a la entrada de las IAG en la creación literaria.

El proceso de esta investigación según lo hemos delineado en los objetivos nos demanda incursionar primero en un criterio estético, por medio de utilizar los conceptos de Autor Modelo y Lector Modelo de Eco. Sobre la base de los resultados de esta primera porción de la investigación,



nos vemos guiados a examinar las consecuencias tanto éticas como epistemológicas de la relación que establecen los textos de las IAG con su lector. De este tipo de repercusiones, la estética no puede brindarnos más que señalamientos semióticos sobre cómo ciertas propuestas estéticas habrán de codificarse en la cultura. Por eso vemos necesario implementar otras perspectivas que permitan ver la cuestión a profundidad.

La verificación que habrá de llevarse más adelante es parte del efecto de esta búsqueda, es decir: se vuelve necesario también entender cómo crea literatura una IAG, cómo es su proceso, y con base a los resultados emitidos, valiéndonos de las herramientas propuestas también por Eco, y que se delinean en el marco teórico, podremos definir mejor esta problemática todavía no explorada de cómo se relaciona un lector con un texto generado mediante una IA.

En síntesis, la forma en que esta investigación contribuye a problematizar las IAG como agentes habilitados para la creación literaria, es a través de la figura del lector. Lo primero fue percibirse de que las IA han alterado un cierto orden antropológico y filosófico al incursionar en prácticamente todas las áreas de lo humano. Y ya que hemos querido observar más allá de los problemas ya evidentes, como el plagio y los derechos de autor, entre otros, recurrimos a los huecos epistemológicos y éticos que se revelan cuando ponemos sobre la mesa el acto de lectura y al lector, pieza clave, junto con la obra y el autor, para el estudio de la literatura.

En la figura 2 hemos incluido un diagrama de Venn para verificar los contactos entre ambas ramas de la filosofía que podrían enfocarse en las problemáticas y repercusiones planteadas. Este diagrama habrá de ampliarse según se construya la investigación. Al ocuparnos de discernir la relación de las obras generadas mediante IAG con sus lectores, nos ocuparemos también de hablar de la influencia en la toma de decisiones y de las problemáticas de los sesgos semánticos, por

ejemplo (véase capítulo III). De momento, continuaremos desarrollando otras partes del apartado teórico y metodológico de este trabajo.

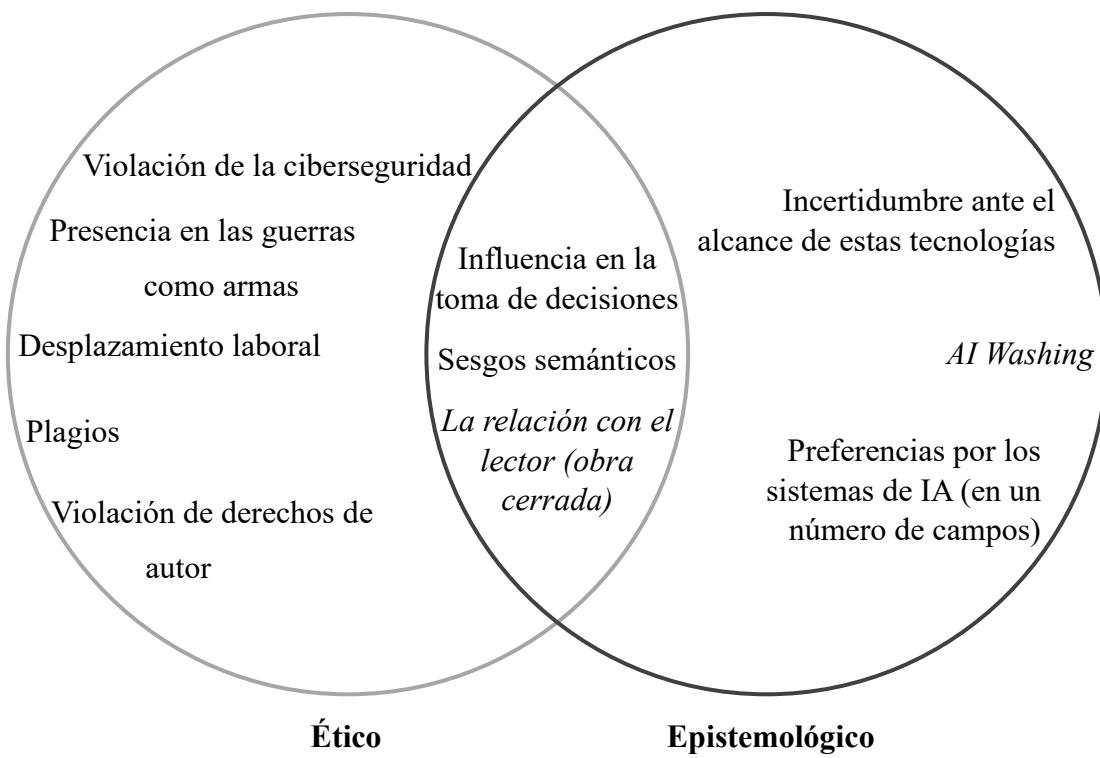


Figura 2



Justificación

Esta propuesta de investigación surge a partir del choque del quehacer literario, como una dimensión de lo humano, con la popularización de las inteligencias artificiales generativas como herramientas para este medio artístico. Queda claro para quienes trabajamos con la literatura, desde su creación o análisis, que la llegada de estas tecnologías tiene el potencial para incidir, de diversas maneras, en un área que concebíamos como humana en su totalidad. No por nada iniciamos el planteamiento del problema con la frase de Elaine Rich, puesto que confirma la llegada de estas tecnologías en cualquier área de lo humano. A raíz de esta observación, constatamos la presencia de varias problemáticas, las cuales consideramos que sientan un antecedente que justifica una investigación de esta naturaleza.

Por otro lado, la propuesta de esta investigación, y de su naturaleza crítica y ética, postula también otra idea: la de que los estudios sobre la incidencia de las IAG en la literatura y las artes se mantengan a la vanguardia, para evitar futuros desencuentros. El recorrido de esta tesis desemboca en la atención de este punto, casi de manera inevitable. Tras todo el apartado dedicado a la relación texto-lector-IA, el capítulo III incursiona en las posibilidades de poner a la vanguardia la teoría y la crítica literarias ante las inteligencias artificiales. En realidad, toda la investigación es un ejercicio dedicado a ello, pero en dicho capítulo, en especial el que refiere a las Humanidades Digitales, tiene como propósito específico proponer que la literatura dé este paso.

Se espera asimismo que los aportes no atiendan únicamente a la literatura respecta. No está fuera de sus objetivos el representar una oportunidad para que programadores y desarrolladores en general de la ciencia de la IA, pudieran obtener una perspectiva más amplia de lo que sucede con estas tecnologías en áreas de las Humanidades. Quizá el problema por adelantado es que estudios de esta índole encuentren un camino hasta esas otras latitudes. Pero el esfuerzo llevado a cabo



también profesa esa intención, como se puede observar en apartados como la metodología y los capítulos II y III en lo particular.

Idealmente, un conjunto de estudios como el que aquí se propone, abre las puertas a contribuir a sentar las bases para la integración de estas tecnologías en la creación artística y reducir el impacto en el trabajo de otros, incluso si es a manera de ampliar el entendimiento de estas tecnologías y así otorgar otra perspectiva al problema de la incertidumbre antes enlistado. Es a través de los objetivos que guían la investigación que se analizarán los modos de trabajar de las IAG, los puntos fuertes y débiles de su funcionamiento, para discernir cómo afectan al trabajo creativo y a establecer un vínculo con el lector. Durante este proceso no se desestimará la consideración del desarrollo de competencias para manejar las IAG. Herramientas que devengan en estimulación y soporte para considerar creaciones más interesantes, para comprender la manera en que analizan obras literarias, por ejemplo. Se puede hacer de una IA un lector ideal de las creaciones literarias o una herramienta de análisis de textos, antiguos, por ejemplo, a través de sus poderosos algoritmos, fenómeno que ya se ha suscitado y que ha levantado admiración por la capacidad no solo de estas tecnologías, sino de utilizarla a ventaja nuestra. En favor de no continuar presenciando problemáticas, las Humanidades en general eventualmente deberán aprovechar su manejo. Y no es sino a través de propuestas como la presente que se obtendrían conclusiones que luego permitan trazar este camino.

Pero también existe, por supuesto, la otra dimensión de la situación, donde no necesariamente tiene que haber una integración de las IA en la literatura y en las Humanidades de forma tajante y total. Ningún fenómeno tecnológico viene dado de este modo. Ahora bien, considerando las protestas sobre plagio, o la incomodidad que generan estas tecnologías ante las creencias de que podrían amenazar con desplazar a cuentistas y guionistas, sería posible desmontar



la supuesta labor creativa de las IAG y puntualizar sus fallas y debilidades, que permitan luego valorar más el trabajo humano. Es decir, desplazar el *AI Washing*, y contribuir a la comprensión del fenómeno de manera extendida y más justa.

En última instancia, justificamos este trabajo con la indagación en el vacío teórico que antes hemos señalado: el de poner al frente la figura del lector y lo que acontece con él en la lectura de un texto generado mediante IA. Justamente porque es un área integral de la literatura, junto con el autor, el libro y la investigación literaria en general, que no ha sido del todo atendida, como se demostrará en los próximos apartados. Y la relación del lector con el texto, el acto de leer una obra literaria, no implica un simple consumo del objeto de inicio a fin. Implica, como se ha señalado apenas superficialmente, cuestiones epistemológicas, por dar un ejemplo, emparentadas también con la ética que acompaña al acto de lectura: la necesidad de mantener la libertad de interpretar una obra. No podemos, pues, pasar por alto de qué manera afectan o no afectan a este proceso tan amplio, dinámico e importante, las creaciones de las IAG, la tecnología en sí misma. De modo que esta investigación deberá abrirse paso en este campo, deliberar sobre esta cuestión prácticamente inexplorada. De hacerlo bien, la teoría literaria se posicionará en el centro de la discusión, dotándola así de un valor de herramienta crítica con la que afrontar también el choque de disciplinas externas a la literatura con esta misma.

Si nos atenemos a que las tecnologías prosigan en su avance, como sin duda lo harán, y no proponemos este tipo de investigaciones, o críticas, lo que podemos esperar es continuar ampliando la lista de problemáticas que devienen del avance de la tecnología sin un entendimiento amplio o regulación. Y si bien un trabajo teórico como este se satisface en su propia calidad de ser teórico, no representa menos que una piedra angular para la indagación de un problema que más adelante pueda dar cuenta ya no de una regulación, sino de un marco ético aplicable, como el que



es propuesto aquí, por áreas especializadas para ellos, como las Humanidades Digitales. Ello debería representar nada menos que un beneficio para las partes involucradas.

Del modo en que se mire, el momento para llevar a cabo estas previsiones teórico-críticas es ahora. El boom de la inteligencia artificial acelera su ritmo. Este hecho incluye consigo una cantidad desorbitante de información, que ha sido dada a través no solo de artículos o tesis, sino también de congresos y ponencias y demás formas hipertextuales. El aprovechamiento de este momento para enfocar la discusión hacia las Humanidades, hacia la filosofía y la literatura, es, por lo tanto, pertinente. Por lo mismo se realizan estudios como el del estado del arte crítico, que recopila y da cuenta de toda esta pléthora de material, con miras a evaluarlo y determinar lo que se investiga actualmente sobre estas tecnologías, acción que, creemos, dota de mayor pertinencia a la investigación en cuestión, y otorga además la ventaja de no sólo sumar este trabajo al cúmulo de todos los otros, sino de realmente considerar un aporte a la materia, del que otros puedan continuar sirviéndose.

Finalmente, los recursos con los que se puede trabajar para una investigación como esta se muestran como óptimos: se tiene acceso a las IAG más avanzadas del momento, en pro de trabajar con ellas y conseguir los resultados más completos y precisos. Por otro lado, es complicado tener un acceso directo a los modelos más avanzados que vayan, no es posible conocer cuándo estarán al alcance de usuarios, o si lo estarán algún día, las versiones más avanzadas. Es necesario, como ya se ha señalado, la expectativa y atención al avance de estas tecnologías. Tal vez esto signifique que el investigador literario, el humanista digital, tenga la inevitable responsabilidad de la vigilancia de estos fenómenos emergentes en la cultura.



Limitaciones y delimitaciones

Una de las delimitaciones clave de este trabajo es la de seleccionar un modelo de IAG de los tantos disponibles con el que llevar a cabo las indagaciones. En nuestro caso será ChatGPT4. Una de las razones fundamentales que acompañan a esta decisión es que nos encontraremos trabajando con base a modelos de prompts¹⁶. Esto se expresa específicamente en el marco metodológico de esta tesis, pero adelantamos que la razón para trabajar con modelos de prompts es para evitar cualquier tipo de obsolescencia en la investigación. De modo que al momento que surja una nueva versión de chatbot¹⁷, el avance de este nuevo modelo respecto al de ChatGPT4 no represente una forma de invalidar la investigación, ya que los métodos son esto, la creación de líneas y estructuras de trabajo que podrían replicarse para nuevos modelos de inteligencias artificiales generativas, considerando, por supuesto, cualquier nuevo agregado que tenga en cuenta sus mejoras y actualizaciones. Pero por la misma naturaleza de los prompts, como los exploramos en el marco metodológico, creemos que las constantes han sido contempladas y trabajadas.

Si nos mantenemos firmes en la decisión de utilizar a ChatGPT4 como representante de las IAG es por la popularidad que ha alcanzado el chatbot en años recientes, ante los chatbots de otras compañías, pero debe tomarse en cuenta también que la tecnología de base es la misma que en otros chatbots. Por otro lado, además del hecho de que trabajaremos con base a modelos replicables, todas las IAG populares del momento están construidas con base a redes neuronales

¹⁶ El significado de prompt es, básicamente, el de instrucción. Véase en marco metodológico una definición más amplia.

¹⁷ “Los chatbots son sistemas que pueden mantener conversaciones extensas con el objetivo de imitar las conversaciones no estructuradas o ‘chats’ característicos de la interacción humana informal” (Jurafsky, Martin, Speech and Language Processing: An Introduction... 243) (“En el original: “Chatbots are systems that can carry on extended conversations with the goal of mimicking the unstructured conversations or ‘chats’ characteristic of informal human-human interaction”).



de tipo Transformador, por lo que las diferencias entre modelos no representan una diferencia crucial, no específicamente para la labor que nos interesa¹⁸.

Reiteramos que esto dota de vigencia a la investigación. La pérdida de tal factor es una amenaza constante para cualquier tesis que se refiera a una tecnología tan sustentable como lo es la IA. Los métodos y enfoques que aquí proponemos mantienen a las futuras indagaciones al respecto con vigencia y con posibilidades de continuar ejecutándose.

Las limitaciones a este trabajo se vuelven más evidentes cuando se pone la mirada sobre el campo de la programación y la informática, áreas que deben ser revisadas según lo plantea la investigación. En los tres capítulos del cuerpo de esta tesis, se ha tenido que incursionar en la comprensión de la IA desde estos campos, y ello ha representado un desafío interdisciplinario. De contar con conocimientos de programación o una noción amplia del funcionamiento de las IA en general, quizás se podrían nutrir ciertos aspectos del marco metodológico y del estado del arte, no obstante, el hecho de no pertenecer el autor de esta tesis agranda las limitaciones al punto de imposibilitar la investigación, de manera que los resultados los consideramos válidos. Únicamente consideramos que el entendimiento de ciertas funciones sería mejor explicado desde las áreas de las ciencias conjuntas encargadas de dar vida a la de la IA.

Una manera de subsanar esta limitante es a través del trabajo conjunto con especialistas del área de ingeniería informática y/o de prompts y en conjunto con humanistas digitales, cuestión que ha sido satisfecha al momento de conformar redes de investigación¹⁹. Por otro lado, argumentamos que no es necesario profesar de un conocimiento a fondo del funcionamiento de la inteligencia

¹⁸ La literatura que se puede encontrar al respecto a mediados de 2025 suele puntualizar comparativas entre áreas diversas. No obstante, por dar un ejemplo, estudios como “A Comparative Analysis of Generative Artificial Intelligence Tools for Natural Language Processing”, de Aamo Iorliam y Joseph Abunimye Ingio, dictaminan que ChatGPT4 frente a Bard, de Google, exhibe mejores aptitudes para la escritura creativa, si bien el último las exhibe para la escritura académica (103).

¹⁹ Las observaciones desde las Humanidades Digitales han sido proporcionadas por la Dra. Elizabeth Takehana, actualmente profesora asociada y directora del Centro de Humanidades Digitales de UTEP.



artificial para cumplir con el objetivo planteado, cuya inclinación es más bien literaria y humanística y no informática, por ejemplo. De cualquier modo, los campos que ameriten una acción intertextual deberán serán, por supuesto, abordados. Así se ha hecho, por citar el ejemplo con más presencia en este trabajo, al momento de hablar de los *embeddings* vectoriales, donde no hace falta especializarse en el aprendizaje a fondo del cálculo vectorial en el contexto del *embedding* de *tokens* de los modelos de lenguaje, pero sí, como hemos hecho, incursionar en la comprensión conceptual del mecanismo de la semántica vectorial y en el cómo una red neuronal Transformador se vale de este para la generación de texto a través de una interfaz de *chatbot*.

Más allá de la consideración de estos artificios y herramientas, no encontramos limitantes que imposibiliten el trabajo a desarrollar a continuación.



Estado del arte crítico: las inteligencias artificiales generativas en la creación literaria

Introducción

En un rango de aproximadamente cuarenta años, las percepciones culturales e investigativas de las inteligencias artificiales creativas, como RACTER, autora del citado *The Policeman's Beard...*, han pasado por varios procesos de codificación y recodificación. No se les trata ya como meras curiosidades de laboratorio: hay un proceso vigente en la cultura por aceptar —o aun por repeler— la presencia extendida de estas tecnologías. Por dar un ejemplo, se está abriendo también la posibilidad de ver a las inteligencias artificiales generativas como autoras de sus propias creaciones, pese a la participación de agentes humanos, ora programadores, ora usuarios de la red con acceso a la variedad de softwares generadores de texto, o *chatbots*. Como ya se dijo, estos fenómenos llevan a la aparición de abundantes estudios, emitidos desde las más distintas áreas y perspectivas. Dada la naturaleza del problema planteado y de la investigación misma que nos proponemos realizar, una manera de revisar estos estudios y pronunciarnos desde el área de la literatura, y desde la semiótica textual como campo de estudio, es revisarlos con una mirada crítica: qué se ha dicho por las distintas autorías de las distintas disciplinas y cómo y bajo qué lente, como un ejercicio enriquecedor de la actualidad de la cuestión de las IA en la literatura, previo a formular nuestros argumentos y en aras de obtener un panorama más allá de lo general sobre las metodologías, críticas, bases teóricas, autorías recurrentes, etc., que acompañan al objeto de estudio —las IA en la creación literaria— en la actualidad.

Dicho esto, hay varias observaciones preliminares que podemos recuperar de entrada. Respecto a las categorías de esta investigación, como los son principalmente los conceptos de Autor Modelo y Lector Modelo y su teoría, es de interés señalar que estas no se abordan en la



literatura disponible de manera directa o tal cual las abordamos aquí. Aun con esto, en la amplia muestra sobre las IA en la creación literaria y sus implicaciones éticas, los conceptos y definiciones recuperados no han de distar demasiado de nuestras pretensiones.

El ejemplo más claro es con la categoría de la creatividad. A esto debemos señalar que esta investigación no pretende permanecer en la línea de la posible incidencia de las IA en el concepto de creatividad. Se toma la postura de que la creatividad se habrá de definir literariamente, como una serie de artificios que evocan al estilo y al idiolecto estético de una obra, es decir, se habrá de definir en términos de las estrategias de lectura y escritura que representan el Autor Modelo y el Lector Modelo. Lo que sucede, pues, en la mayoría de los artículos, es que estas categorías no se abordan, y se permanece ante todo con el concepto de la creatividad. Sin embargo, esto no perjudica la validez de la investigación documental. Todo lo contrario: representa ya un primer hallazgo desde el cual empezar a esbozar conclusiones, como podría ser la falta de estos estudios de considerar los términos de la teoría literaria aun cuando se está trabajando con obras literarias, por ejemplo.

Asimismo, puntualizamos de nuevo la observación de que los manuales de regulación para inteligencias artificiales propuestos por organismos especializados, como la UNESCO, la UE, la Universidad de Montreal, entre otros, expresan una preocupación más ávida por las cuestiones de ciberseguridad y el desplazamiento de los empleos, omitiendo la integración de las disciplinas creativas en la discusión. Es por eso que a pesar de que por supuesto que integran parte de la discusión, no se incluyen como tal en este estado del arte crítico, guardándose su uso específico para el capítulo tres.

Siendo estos los únicos antecedentes, delineamos a continuación la metodología empleada para este estado del arte crítico.



Metodología del estado del arte

El estado del arte crítico forma parte de la metodología de esta tesis para la investigación documental, la cual es de corte hermenéutico. Para la elaboración de este, nos hemos valido del apoyo de una matriz metodológica del estado del arte elaborada en Excel: en esta se creó un corpus de 49 fuentes, incluidos libros, artículos y una entrevista en formato de video. Tras una revisión de estos textos bajo los criterios establecidos, y tomando en cuenta las categorías de esta tesis, se eligió la cantidad de 28 fuentes como las más pertinentes para someter a revisión en esta investigación documental.

El primer lineamiento a tener en cuenta para la reunión de material es establecer un límite diacrónico: se considera lo que se ha escrito y reflexionado en los últimos diez años sobre el objeto de estudio. La restricción de años en la búsqueda de los estudios no viene dada por mero protocolo: el tema de las inteligencias artificiales ha venido explotando en un plano cuantitativo sobre todo en los últimos cinco años, hecho adjudicable especialmente a la compañía *OpenAI*, pero también a los modelos Transformadores, de los cuales la compañía Google ha sido el pionero en 2017. Sin embargo, para conformar un estado del arte que en verdad abra el panorama de la investigación sobre estas tecnologías, se vuelve necesario considerar algunas excepciones a este rango que se consideren pertinentes. Las razones para no extenderse más de los diez años son sensiblemente distintas, y es que la aparición de tanto material académico e investigativo de los últimos años dificultaría la tarea de investigación, saturándola. Más fructífero ha sido establecer criterios puntuales de búsqueda con las categorías que se manejan en esta investigación (semiótica textual, Autor Modelo, Lector Modelo, lector, estilo-literario, creatividad, ética de las IA, etc.) y así



establecer una visión global, pero que no sobrepase a estas categorías, del estado del arte, lo cual debería llevarnos a conclusiones igualmente satisfactorias.

Finalmente, el añadido crítico de este estado de arte nos ha llevado a analizar los documentos recopilados en función de qué es lo que se ha escrito, cómo y desde qué áreas y aun qué hace falta por ser investigado sobre las inteligencias artificiales en las áreas de la producción creativa. Con la metodología base, las categorías de esta investigación por delante y esta resolución de observación crítica, se ha resuelto a establecer cuatro apartados: el primero y más pertinente a esta tesis, delinea los resultados sobre la creación literaria de las IAG observada desde la teoría semiótica de Umberto Eco, que trabaja con base a la relación del texto con su lector; el segundo expone el trabajo creativo como tal de las IA, donde se perciben cuestiones del trabajo de imitación del estilo de un escritor o un género, y del que surgen también intentos de crítica a dichos trabajos; el tercero va sobre los encuentros de las IA con conceptos y metodologías de la teoría literaria para evaluar su forma de crear y dar algunos fallos al respecto también; y, por último, el cuarto exhibe lo que se ha dicho sobre las implicaciones éticas que de esta conjunción de campos se puedan prever o identificar.

1. La teoría de la semiótica textual de Umberto Eco aplicada a las creaciones de las IAG

De los cinco artículos seleccionados para este segmento del estado del arte crítico, solamente uno aborda el tema de manera cercana a como lo hemos planteado en los objetivos. El resto aborda la teoría semiótica de Eco tangencialmente. Su preocupación parece ser más bien la de generar modelos semióticos-narrativos que puedan luego integrarse en las inteligencias artificiales generativas, con la finalidad de mejorar la calidad de las historias producidas. Este



acontecimiento no es, de hecho, ajeno a los siguientes dos apartados de este estado del arte, es, de hecho, una “preocupación” extendida esto de la búsqueda de modelos o métodos cuando se trata a las IAG desde una perspectiva literaria. Lo que nos interesa de este grupo de artículos en conjunto es a) la extensión de estas propuestas de modelos narrativos, y b) el uso de la semiótica textual como un punto de anclaje para cumplir con el inciso a).

Especialmente la intervención titulada “*Start Making Sense: Sociosemiotic contributions to the understanding of Generative Artificial Intelligences*”, de Carlos Scolari, trabaja de la mano de la semiótica textual de Umberto Eco y cómo esta incluso ha tenido sus acercamientos con las inteligencias artificiales a través del concepto lingüístico de *frame*, propuesto por Marvin Minski para su uso desde la IA en el contexto de la comprensión lingüístico-narrativa (“Frames”, 1). Por otro lado, no hace mención de los conceptos principales de la semiótica textual, como lo son su lector y el texto, sino que delinea más bien de manera general las contribuciones de la semiótica a esta ciencia, especialmente de la *sociosemiosis*, como un apartado pragmático y contextual que sitúa la semiótica como una herramienta de lo social. Este mismo proceder tienen los artículos de “*An AI-Powered approach to the semiotic reconstruction of narratives*”, de 2024, y “*n Walks in the Fictional Woods*”, de 2023. Este último en especial propone un modelo visual de narrativa cuya base teórica descansa en su totalidad en la obra de Eco *Seis paseos por los bosques narrativos*, cuya propuesta es narratológica, pero también con una consideración especial a las inferencias que de un texto puede elaborar su lector. Su postulación es interesante en tanto que representa un modelo concreto que trabaja sobre una base sólida como lo es la semiótica textual. No obstante la calidad de su propuesta, este, como los otros también, se hayan todavía en condiciones experimentales. El artículo “*An AI-Powered approach...*” por su parte trabaja más bien desde un área especializada que se denomina “narratología computacional”, de la que han hablado también



otras autorías. Si bien este campo no es fundamental para nuestro estudio, registrar su existencia determina la evidente preocupación que existe por la integración de modelos narrativos, modelos semióticos en general, como una herramienta para las IAG.

La duda que nos asalta ante este primer “síntoma” es respecto a la necesidad de ver cumplido de este cometido. Los artículos no expresan esto con claridad. Incluso en “*n Walks...*”, los autores hacen un llamado casi ético hacia el final de su discusión, lo cual es de subrayarse, dado que eso evoca una similitud cuando menos estructural con nuestra presente investigación, y porque parece emitir un doble mensaje, donde por un lado se busca la mejora de las IAG en su calidad y capacidades narrativas, y por el otro se hace un cuestionamiento filosófico sobre el papel de las máquinas para nuestra creatividad:

A medida que avanzamos en nuestras capacidades tecnológicas, estas historias sirven como recordatorio de la necesidad de cautela, responsabilidad y respeto por todas las formas de inteligencia, ya sean biológicas o artificiales. ¿Y es aquí donde trazamos el límite? ¿Todas nuestras herramientas merecen amor? ¿Qué pasa con la inteligencia ficticia, si alguna vez descubrimos que existe en alguna forma?²⁰ (Schetinger, Victor, Di Bartolomeo, Sara, de Lima, Edirlei et al. 7)

De este modo, el porqué es de considerarse la forma de inteligencia artificial del mismo modo que la biológica, no se establece con claridad. Incluso se pone en duda la existencia de la inteligencia ficticia en la misma línea. Así, pues, la respuesta a por qué habría de dotar de estas capacidades a estas tecnologías no encuentra ni siquiera su pregunta planteada. Sin embargo, en este punto la

²⁰ En el original: *As we continue to advance in our technological capabilities, these stories serve as a reminder of the need for caution, responsibility, and respect for all forms of intelligence, whether biological or artificial. And is it here that we draw the line? Do all of our tools deserve love? What about fictional intelligence, if we ever find out it exists in some form?*



reflexión se aleja lo suficiente de la semiótica textual, por lo que abordaremos el siguiente documento.

En “*AI Literature: the disenchantment of word and the enchantment of code*”, el uso de la semiótica textual se determina en función de que al no poseer un autor “real”, la clave para la interpretación de sus textos reside meramente en una teoría centrada en la recepción, es decir, en el lector. Sin embargo, el interés del artículo no se centra en establecer esta relación, sino en las consecuencias mismas de las IA como productoras de literatura. El “encanto” por el código profesa una exaltación de la tecnología que vale la pena replantearse. No extraña, pues, que dentro de sus referencias el artículo cite a Adorno y Horkheimer e incluso a Benjamin Walter, por obvias razones. En cuanto a la teoría de la recepción, dada también desde una semiótica textual, no hay un esbozo metodológico puntual ni sistemático al respecto.

El artículo que, por lo tanto, captura la mayor parte de nuestra atención es el de “*Echoes of Intelligence: Textual Interpretation and Large Language Models*”, de Alvaro Videla, 2023. No sólo en este tienen relevancia los conceptos de Autor Modelo y Lector Modelo, propios de Eco, sino que también una parte de la argumentación nos interesa. En general, el artículo remite a las categorías de nuestra investigación, y por ello analizamos sus conclusiones. Las razones que las diferencian son importantes, sin embargo, y podemos partir desde ellas.

Lo que nos interesa en este apartado del E.A.C es el tratamiento de la relación texto-lector explorada a través de la semiótica textual²¹. Y justamente lo que Videla postula como su hueco teórico es que una IAG²² no posee un propósito para escribir sus textos, dejando ausente así la

²¹ En el apartado tres de este E.A.C la semiótica deja de representar un criterio de búsqueda y se acude a Leia Henricksen, quien propone como clave de lectura de textos de IA la hermenéutica.

²² El autor utiliza el término LLM, “*Large Language Models*”, para referirse a las IA con la capacidad de generar texto. Para efectos de esta parte de la investigación, nos permitiremos considerar el término como sinónimo de IAG. Las diferencias y la relevancia de estas se explicitarán más adelante en el cuerpo de la tesis. De momento podemos prescindir de separar ambos términos.



parte pragmática que acompaña cada texto, hecho el cual podría representar una brecha importante al momento de que un lector aborde el texto de una IAG, texto que el autor llama *sintético*, para diferenciarlo del texto de un humano que sí cuenta con una pragmática, con un rostro legible, por así decirlo.

Una diferencia entre texto escrito por un humano y texto sintético generado por un LLM²³ es que el primero es producido por un autor con intenciones ... un autor toma muchas decisiones para asegurar que el receptor semántico de su mensaje encaje las características semánticas-estadísticas de aquellos sus receptores (41).

Esta ausencia de intenciones conlleva la ausencia de visión semántica y pragmática de un texto y, por lo tanto, asume Videla, el texto se “desconectaría de su lector”, lo cual acarrearía otro tipo de consecuencias, cuya causa puede ser enunciada como el desconocimiento de que la autoría del texto era una inteligencia artificial. Es decir, un señalamiento estético.

La solución radica para Videla en construir una competencia, a la manera de la estrategia que representa el Autor Modelo²⁴, que indique que un texto sintético es eso: la construcción de una inteligencia artificial. Esto se puede lograr desde varias maneras, pero el postulado es ese, que el lector de un texto sintético reciba las indicaciones suficientes para comprender que está leyendo algo generado mediante una IAG. Y esto, asegura el autor, contribuiría a reducir las brechas entre el lector y el texto, naturalmente causadas por la actuación de la IA. Esta argumentación la respalda a través de la propuesta tanto de Eco como la de Wittgenstein con los juegos del lenguaje. “Entonces, ¿cuáles son las reglas de un juego de lenguaje con un LLM? ¿Cuáles son las intenciones

²³ En español se suele traducir como Modelo de Lenguaje de Gran Tamaño o Modelo Extenso de Lenguaje.

²⁴ Para la definición exhaustiva de Autor y Lector modelos véase el apartado homónimo en el marco teórico. De momento basta con enunciar que estos conceptos representan estrategias de construcción textual: el Autor Modelo de un texto es construido por su lector en un intento de figurarse, de interpretar qué está queriendo comunicarle el texto, mientras que el Lector Modelo es una estrategia prevista por el autor empírico, en su intento de prever la forma del posible lector de su texto, competencia que él mismo construye en tanto que urde el texto.



del texto sintético?” (42). Esta propuesta contribuiría a evitar el también comentado por él efecto ELIZA, donde un lector se muestra susceptible a llenar esos huecos en un texto con información innecesaria y hasta inadecuada solo por el hecho de saber que el texto fue generado por una computadora (42-43), entre otras consecuencias que no distan mucho de lo que antes hemos expuesto como *AI Washing*.

Las observaciones de Videla son agudas, pero el autor comete un error al asumir, de entrada, que una IAG se opera por sí misma, lo cual lo lleva a argumentar que un concepto como el Autor Modelo no es aplicable para estas tecnologías. Basta con recurrir al concepto de *prompt*, que es la clave para la comunicación entre un usuario humano y la máquina, para poder pensar la cuestión en términos de autorías empíricas y Modelo. Si bien el autor señala hacia el final de su artículo que suele haber dos operaciones llevadas a cabo por la IAG, una de las cuales involucra la cooperación de un usuario humano, su distinción no parece lo suficientemente fundamentada para su propuesta. A fin de cuentas, lo que plantea es que ambos tipos de operaciones tienen bases de datos diferentes para conformar sus respuestas, una es una base de datos más general mientras que la otra viene dada por la comunicación con el usuario. Como dijimos, esta distinción puede ser soslayada en cuanto que una IAG no posee la autonomía para generar texto literario, o de cualquier tipo. Necesariamente debe contar con la operación de un humano para ella misma operar luego, de cualquier manera. Aun y cuando la distinción fuera crucial, el hecho de que Videla la plantee hacia el final, justo antes del apartado de conclusiones, debilita la posibilidad de argumentar sobre el hecho.

La propuesta que nosotros esbozamos en la investigación es que un texto sintético —en la terminología de Videla— sí puede contar con un Autor Modelo, y aun con un autor empírico. De momento, lo que rescatamos de su propuesta es que al argumentar la necesidad de una manera de



informar al lector que está ante un texto sintético, está proponiendo una necesidad de *transparencia*. Transparencia que indicaría que el texto está escrito por una IAG con X e Y características que el lector no debería soslayar. Algunas posibles soluciones: transparencia, a fin de cuentas. Este punto nos parece pertinente incluso para el cuerpo de la tesis.

Como se ha observado, este artículo remite a los conceptos de la semiótica textual de Umberto Eco y plantea preguntas pertinentes sobre las condiciones en que un lector se enfrenta con un texto producido por una IAG, que es con precisión lo que rastreábamos en este apartado. Si por un lado hemos sido críticos con el texto de Videla es porque creemos que el hueco teórico se inclina por otra dirección, hecho que no invalida lo pertinente de su investigación, a la cual habremos de volver más adelante. En términos de la conformación de un estado del arte crítico, el resto de los artículos examinados formula muy bien el panorama que existe entre ambos campos, los puntos encontrados y las brechas que hay entre ambos, que no expresan otra cosa que la posibilidad de explorarlas teóricamente.

2. Inteligencias artificiales generativas y la creatividad: posibilidades y problemáticas en la literatura

Este apartado integra artículos que conjuntan la visión de la capacidad creativa de las inteligencias artificiales y su participación en coacción con la creatividad humana, y más aún: algunos consideran extenderse hasta la ética para abordar dicho tema, dado que existe la posibilidad de que esta creatividad venga dada como una manera de reemplazar la del usuario humano. La creatividad como concepto parece ser el que más relevancia tiene antes que cualquier concepto meramente literario que explique las particularidades de las obras. Es decir, no se da un tratamiento en términos de estilo o de poética. En las más de las veces, el concepto, o la capacidad



para mostrarse creativas, aparece abordado como un algo demasiado general, incluso algo que ya está dado y no hace falta demostrar las cualidades de cómo algo puede ser creativo y respecto a qué, anulando, por lo tanto, la aparición de cualquier base teórica sólida que la defina. Otro conjunto de artículos parece más obstinado en señalar los huecos que las IA tienen para ejecutar creaciones que valgan la pena o sean, por tanto, creativas, pero lejos de aproximarse a una propuesta de un modelo, a la manera en que se observó en el apartado anterior, sus señalamientos rayan en propuestas demasiado preliminares. Esta forma de proceder es lo que antes hemos aludido: ¿en beneficio de qué área se esbozan estas propuestas? Por más preliminares que sean, es una de las dudas principales de este estado del arte crítico: de integrarse estos modelos narrativos, semióticos, o de hacer caso de las propuestas que se verán a continuación, ¿cuál es el propósito? Incluso si fuera por motivos meramente experimentales, motivos válidos, por supuesto, la pertinencia de los artículos demandaría que se enunciara, mas no se lo enuncia, dejando a la interpretación los posibles porqués.

En el grupo de artículos que argumentan sobre cómo las IAG podrían representar un impulso para la creatividad encontramos uno de corte cuantitativo y otro cualitativo. El primero, “*Generative artificial intelligence enhances creativity but reduces the diversity of novel content*”, y el segundo, “*Literature in the Age of Artificial Intelligence: A Preliminary Study on the Big Language Model AI*”. Ambos artículos, como se ha señalado ya, presentan una noción de creatividad endeble. El caso está dado en el balance de los resultados del estudio cualitativo: “Encontramos que tener acceso a GenAI aumenta causalmente la novedad y la utilidad promedio (dos dimensiones de la creatividad frecuentemente estudiadas) en relación con los escritores humanos por sí solos”²⁵ (Anil Doshi, Oliver Hauser 10). De modo que el resto de las conclusiones

²⁵ Original: *We find that having access to GenAI causally increases the average novelty and usefulness—two frequently studied dimensions of creativity—relative to human writers on their own.*



podrían ser problemáticas. Lo mismo puede decirse del instrumento de medición con que se llevó a cabo la investigación. No hace falta, sin embargo, ahondar en la creatividad como concepto para leer exitosamente una buena parte de las conclusiones. En términos generales se alude a que los escritores “menos creativos” son quienes se beneficiarían de las IAG (11), en tanto que incluso un escritor más experimentado, podría ver mermada su creatividad ya que una IAG sometería el escrito a una restricción de “la variabilidad de las propias ideas del escritor desde el principio, inhibiendo la creatividad abierta” (2).

De cualquier modo los autores dan a entender que existen posibilidades para llevar a cabo ese impulso de la creatividad, pero no es su intención describir cómo se darían estas posibilidades.

El segundo artículo, de la mano de Yan Hu, expone una serie de posibilidades que parten desde la siguiente premisa: “A medida que estos modelos imitan cada vez más el estilo y el vocabulario de la escritura humana, tienen el potencial de revolucionar todos los aspectos de la escritura creativa en el futuro”²⁶ (1782). La imitación del estilo y del vocabulario como ejes de la escritura humana no son abordados de manera más amplia, y a partir de aquí se lanza el autor a la argumentación. Algunos de los puntos que toca los consideramos válidos y pertinentes, ya que las preguntas planteadas llevan a considerar, cuando menos, otras preguntas. Por otro lado, se presentan algunos “síntomas” compartidos entre los documentos.

Parece ser que a) los artículos aquí propuestos (los ya abordados y los que resta abordar) toman de soslayo a la creatividad, y por lo tanto, b) no argumentan a profundidad las consecuencias de las exhibiciones creativas de las IA, como la incidencia en la creatividad humana, etc. Como hipótesis, creemos que un estudio fructífero sobre la creatividad en las IAG debería contar con una base teórica que defina no sólo la creatividad, sino la creatividad abordada desde la informática.

²⁶ Original: *As these models increasingly mimic the style and vocabulary of human writing, they have the potential to revolutionize every aspect of creative writing in the future.*



Para este efecto, volvemos a nombrar a Margaret Boden, cuyo artículo cuenta con una aproximación más ventajosa al problema, si bien para el tiempo en que fue publicado, todavía quedaban muchos huecos en la teoría. Este es el principal fallo, parece ser, cuando se habla de las posibilidades creativas de las IA. Lo que proponemos en esta tesis es establecer las categorías desde la literatura de manera que, a través de estas, y con el respaldo de otros conceptos literarios, se pueda llevar a cabo la enunciación de lo que hacen las máquinas que es estéticamente interesante o no. De otro modo, la categoría de la creatividad, y las consecuencias que los autores exponen, admiten la duda del lector.

Para continuar con el artículo anterior, debemos actuar del mismo modo que antes y asumir que entendemos que la creatividad está en juego, para bien o para mal, cuando se habla de inteligencias artificiales generativas, si bien no se nos explica el cómo. El autor procede, pues, por dos vías: la una se muestra fatalista, irrumpiendo con el concepto de “*Creative Destruction*”, y se suma a la visión sociológica, y aun ética, de las IAG como una posible amenaza para la creatividad, estirando las posibilidades a un grado más amplio.

La “destrucción creativa” se refiere al proceso en el que el crecimiento económico y la innovación tecnológica promueven la transformación industrial, lo que a menudo lleva a la eliminación de industrias antiguas y al surgimiento de nuevas industrias ... la “destrucción creativa” impulsada por nuevas tecnologías dará forma a una nueva cultura sobre la base de promover la mejora de la eficiencia industrial²⁷ (1782-1783).

²⁷ Original: ‘*Creative destruction*’ refers to the process in which economic growth and technological innovation promote industrial transformation, often leading to the elimination of old industries and the emergence of new industries ... ‘creative destruction’ driven by new technologies will shape a new culture on the basis of promoting the improvement of industrial efficiency.



A través de esta noción, señala Yan Hu que amplios cambios en la cultura se pueden vislumbrar, sin ahondar en el cómo. De manera que se problematiza la existencia de máquinas creativas, más allá de la máquina misma.

Pero esta argumentación conserva su validez cuando se enfoca la mirada en la literatura. De modo que, sin perder de vista desde dónde parten los postulados del autor, quizá el mayor acierto de su artículo sea este señalamiento:

Cuando los creadores humanos crean literatura y contenido, siempre están influenciados por los antecedentes de su propia época y las corrientes estéticas y filosóficas generales de la sociedad. Sin embargo, en el caso de las IA de modelos de lenguaje grandes, sus modelos centrales se entrena con datos de texto que los humanos han podido encontrar durante miles de años. Para ellas, el trasfondo de la época y la estética y filosofía únicas de la época no constituyen el significado de “pensar”. Para ellas, la epistemología y la ontología de las personas en cualquier época solo contienen algunas diferencias en los datos, y no afectan su estilo habitual de creación de contenido²⁸ (1785).

Estos señalamientos son todavía más interesantes desde un punto de vista literario y filosófico, y es uno de los derroteros a que se abre la presente tesis en sus capítulos. No obstante, cuando entendemos lo endeble de la fundamentación con la categoría de la creatividad, de la “destrucción de la creatividad”, permanece en nosotros la duda de la verdadera relevancia de estas nociones.

Pero finalmente, evitando caer en el maniqueísmo al que recurren otras autorías, señala que las IAG, por su disposición y fácil acceso, podrían representar también un apoyo para la

²⁸ Original: *When human creators create literature and content, they are always influenced by the background of their own era and the general aesthetic and philosophical views of society. However, for large language model AI, their core models are trained by text data that humans have been able to find for thousands of years. For them, the background of the times and the unique aesthetics and philosophy of the times do not constitute the meaning of “thinking” for them. For them, the epistemology and ontology of people in any era only have some differences in data, and it does not affect their usual style of content creation.*



creatividad. Yu Han dibuja algunos escenarios con los que es difícil estar de acuerdo, no porque se muestren imposibles, sino porque la teoría que opera tras esos escenarios brilla por su ausencia, y quedan como meras menciones de posibilidades que no sabemos si algún día podrían llevarse a cabo y cómo. Nos referimos, por supuesto, a las menciones que hace sobre las personas con discapacidades físicas, y a la visión de comunión hacia el final de la cita. Por lo demás, las menciones a la absorción del estilo del escritor humano, y al impacto al concepto de autoría, a simple vista son posibilidades mejor concertadas.

...el uso de la IA en la creación literaria puede ayudar a los autores humanos y a la IA entrenada por ella a formar un sistema creativo único... Cualquiera puede llamar directamente al modelo de datos de la IA para que se sirva a sí mismo. El creador puede digitalizar su corpus escrito previamente y subirlo a una base de datos. A continuación, dejar que la IA lo imite y genere contenido literario único para él que se adapte a sus características creativas... El desarrollo de este fenómeno en el futuro puede tener un impacto en la teoría del concepto tradicional de autoría en el futuro. Este impacto puede extenderse a otras áreas de producción de contenido, como la curaduría y la publicación ... La IA de Grandes Modelos de Lenguaje, también puede impulsar a más personas a unirse a las filas de los creadores ... Este avance brindará un camino creativo a las personas con discapacidades físicas. Además, debido a las importantes diferencias entre los diferentes idiomas, a menudo un excelente autor sólo puede escribir con fluidez en su lengua materna ... promoviendo así la cooperación y los intercambios de culturas y disciplinas en todo el mundo²⁹ (1785-1786).

²⁹ En el original: *...the use of AI in literary creation may help human authors and AI trained by it form a unique creative system ... Anyone can directly call the AI data model to serve themselves. The creator can digitize his previously written corpus and upload it to a database. Next, let AI imitate him and generate unique literary content for him that suits his creative characteristics ... The development of this phenomenon in the future may have an impact*



Artículos como este en especial condensan muy bien lo que el resto de artículos en la lista esbozan en cuanto a la creatividad de las IAG en la literatura. Estos aducen también algunas de las problemáticas que derivan de las IAG haciendo presencia en este campo, y también adolecen de detallar entre sus párrafos lo que consideran por creatividad.

“Hacer que las estúpidas máquinas canten”, construye un aparato crítico de la mano de nombres como Walter Benjamin y la teórica literaria Katherine Hayles, y en conjunto con una lectura de los algoritmos como “mecanismos artísticos” (Rico 113) crea un balance entre la asimilación de las capacidades creativas de las IAG y el cuestionamiento de este hecho. Aborda una propuesta categórica que define como e-literatura, es decir, toda creación literaria, cuento, novela o poema, que excluya el libro impreso y que se valga de la tecnología digital para su creación y recepción (112). En torno a esto analiza la acción y el alcance de varias inteligencias artificiales como puentes de la generación de esta e-literatura, empezando por la comentada en esta tesis RACTER, el chatbot Eliza y los modelos Transformadores. Aun con todo, y no obstante la solidez presentada en cuanto a los términos teóricos, el fallo de Rico se inclina ligeramente a favor del uso de las IAG para la creación literaria, sin extenderse cuando menos en un listado de las razones para ello, y, por supuesto, sin involucrar las consecuencias derivadas, a la manera del artículo de Hu.

...la generación de textos mediante inteligencia artificial podría acarrear la posibilidad de ampliar los horizontes de la escritura, sin que esto implique necesariamente una aniquilación de las formas de composición o de lectura más tradicionales. Tampoco puede

on the theory of the traditional concept of authorship in the future. This impact may extend to other areas of content production, such as curation and publishing ... Big language model AI may also drive more people to join the ranks of creators ... This advancement will bring a creative path to those with physical disabilities. In addition, due to the significant differences between different languages, often an excellent author can only write proficiently in his native language ... thereby promoting cooperation and exchanges of cultures and disciplines around the world.



ignorarse que la generación de estos textos todavía requiere una instrucción y una pregunta ni que su reconocimiento como escritura o su final adecuación a un contexto particular todavía dependen de una evaluación lectora, que, como se sabe, no es nunca del todo pasiva (117).

“*AI as Author: Bridging the Gap Between Machine Learning and Literary Theory*”, comparte el punto de vista de Hu: argumenta que las inteligencias artificiales y los algoritmos, “podrían transformar potencialmente el modo en que vendemos, leemos y reseñamos libros” (Heerden, Bas 175)³⁰.

Es interesante llevar la mirada a las dos autorías del artículo, van Heerden, que se encuentra en el área de literatura comparada, y Bas, del área de ingeniería computacional. La estrecha comunicación de ambas disciplinas develaría un material más puntual y acertado en cuanto a sus revisiones y propuestas. Y en efecto: el artículo esboza un ejercicio de redacción de poemas utilizando la IAG *Deep-Speare*, un programa diseñado específicamente para la creación de poemas en el estilo de Shakespeare, revisa la teoría literaria desde algunos de sus conceptos elementales y termina haciendo un análisis que apunta a un argumento central, vehículo de todo el artículo. Propone entonces, como hicieron los artículos del primer apartado, el enfoque en la teoría literaria para mejorar la calidad de la “literatura generada por computadora” (178). Con qué fin o en beneficio de qué área, no se expresa. La posibilidad únicamente se deja abierta: “Los problemas que actualmente enfrentan los escritos generados por máquinas podrían abordarse de manera más efectiva aplicando comprensiones teóricas literarias de creatividad, originalidad, ambigüedad y emoción, entre otras”³¹ (178). Estos problemas son, por supuesto, la falta o el exceso de lo que se

³⁰ En el original: *could potentially transform the way we sell, read and review books.*

³¹ En el original: *To improve the quality of computer-generated literature, we suggest combining tools and insights from various text-centred (rather than biographical, cultural or socio-historical) approaches in literary theory. Involving true expertise on what literature is and how it works would strengthen current research on learning-based*



enuncia: creatividad, originalidad, ambigüedad y emoción. ¿Cómo lograría la implementación de conceptos de la teoría literaria la aprehensión semántica y el uso pragmático de estos conceptos? Tampoco se enuncia. Quizá uno de los problemas fundamentales devenga del reduccionismo evidente en su marco teórico. Nos referimos a la definición operativa que se maneja de literatura: “Para este trabajo, definimos literatura simple y tradicionalmente como escritura imaginativa o inventada, como ficción (Eagleton, 2011)³²” (176). Cita a otras autorías, evocando incluso puntos cercanos a la función poética del lenguaje de Jakobson en un pie de página y líneas más adelante (176, 179). No obstante, el problema de este primer esbozo es que no termina de explicar lo que es la literatura. Estos reduccionismos tan recurrentes en las categorías y conceptos operativos nos remiten a los índices antes marcados, en especial al b), donde argumentamos que las consecuencias de máquinas creativas no logran preverse del todo cuando esta creatividad no es adecuadamente definida. De definirse esta creatividad, muchas conclusiones se podrían extraer, como hemos hecho, bajo nuestro aparato teórico, en los capítulos I y II.

Como hicimos antes, ya solo rescataremos el mayor acierto del artículo, que llega hacia el final de este, aunque en algunos puntos parezca un ejercicio más bien retórico. El primer punto destaca al menos la necesidad de resignificar, tal vez, el concepto de autoría.

¿Veremos el aumento de las editoriales de IA o de los departamentos de IA dentro de las editoriales? ¿Cómo operarán y qué desafíos legales y éticos enfrentarían (por ejemplo, plagios potenciales y violación de los derechos de autor)? Van der Weel (2015) destaca la

text generation systems. The issues currently faced in machine-generated writing could be addressed more effectively by applying literary theoretical understandings of creativity, originality, ambiguity and emotion, among others.

³² Original: *For this work, we define literature simply and traditionally as imaginative or invented writing, like fiction (Eagleton, 2011)*



necesidad de nuevas definiciones de autoría y derechos de propiedad intelectual en relación con el desarrollo tecnológico³³ (181).

Esta segunda cita es la que parece operar todavía más a nivel retórico, pero al menos proporciona otra base teórica desde la cual se desmonta la noción de creatividad y originalidad de las máquinas.

¿Qué impacto tendrá la literatura de las IA en las definiciones de originalidad y creatividad?

Según Klebanov y Madnani (2020), actualmente no existe ningún sistema de puntuación operativo que dé prioridad a la originalidad en el texto generado y “una vez que se puedan medir con éxito varios indicadores de originalidad, puede ser necesario trabajar más para incorporar estas mediciones en los ecosistemas de puntuación”. Los rasgos y criterios de medición de la originalidad aún no se han determinado en la lingüística computacional (Klebanov y Madnani, 2020) y, fuera de este contexto, se consideran inalcanzables (Gruber, 1988). En un contexto literario, Gross (1997) explica que los intentos de categorización podrían fallar en hacer justicia a la singularidad y el poder de la poesía y, por lo tanto, requieren una gran destreza³⁴ (182).

Es necesario volver a anotar la conclusión de que las observaciones sobre las categorías en torno a la capacidad creativa-literaria de las IAG no son detalladas, y está latente el peligro de caer en exageraciones, o incluso en la misma corriente que generan fenómenos como el propio *AI*

³³ Original: *Will we see the rise of AI publishing houses or AI departments within publishing houses? How would these operate and what legal and ethical challenges would they face (for example, potential plagiarism and copyright infringement)? Van der Weel (2015) highlights the need for new definitions of authorship and intellectual property rights concerning technological development.*

³⁴ Original: *What impact will AI literature have on definitions of originality and creativity? According to Klebanov and Madnani (2020), there is currently no operational scoring system that prioritises originality in generated text, and ‘once various indicators of originality can be successfully measured, additional work may be necessary to incorporate these measurements into scoring ecosystems’. Traits and measurement criteria of originality have yet to be determined in computational linguistics (Klebanov & Madnani, 2020) and have, outside this context, been considered unachievable (Gruber; 1988). In a literary context, Gross (1997) explains that attempts at categorization may fail to do justice to the uniqueness and power of poetry and, therefore, require great dexterity.*



Washing. Este hecho conforma una parte importante y reveladora de un estado del arte sobre las IAG y la literatura, pero en el siguiente apartado habrán de hacerse notar artículos con un enfoque mucho más perceptible, ya que se trata de autorías que hablan, en su mayoría, desde el interior de la literatura, y sus propuestas y críticas pasan a ser menos ambiguas.

3. Aproximaciones a las IAG desde la literatura

En los artículos que aquí se reúnen, la creatividad, de cierta manera, se da por sentado. Lo que llama la atención es ver la aproximación literaria a las creaciones de las IAG y a sus capacidades, como la de imitar el estilo de un autor, por dar un ejemplo. Ciertos puntos habrán de tocarse, sin dudas, con los apartados anteriores, de modo que, aunque se dé por sentada, aquí también se toca el tema de la creatividad de las IA, y no faltan tampoco las propuestas sobre cómo podrían mejorar sus competencias literarias, pero los enfoques centrales cambian: las autorías de este apartado nos revelan lo que desde áreas específicas de la teoría literaria se ha dicho sobre las inteligencias artificiales en la literatura, hecho que revela a su vez qué tan pertinentes resultan estos estudios de cara a estas tecnologías. Hemos mencionado en el corpus problemático de esta investigación que la literatura debería mantenerse a la vanguardia frente a la evolución de las IA, y esto es lo que evaluaremos más de cerca en este apartado específico.

El libro de Jorge Carrión y otros, *Los campos electromagnéticos*, abre con una propuesta entre lo optimista y lo experimental. La idea que profesa es la de “invitar” a participar de lo literario a las inteligencias artificiales. En esta noción lo acompañan, aunque sin profesar el mismo optimismo, “*Synthetic Literature. Writing Science Fiction in a Co-Creative Process*” y “*Simulating horror literature by H.P. Lovecraft with ChatGPT*”. El primero de este par acarrea el propósito de estimular la creación, la *co-creación*, por medio de un programa especializado para la ciencia



ficción, en aras de constituir un punto de apoyo para el escritor dedicado a este género. “Nuestro objetivo es crear un entorno estimulante que fomente la cocreación: idealmente, la máquina debería generar sugerencias valiosas, en las que el autor mantenga un interés importante dentro del proceso creativo”³⁵ (Manjacavas et al. 1). El segundo estudio expone las posibilidades de que ChatGPT emule el estilo de Lovecraft, para convertirlo luego en una herramienta de escritura intertextual: de réplica o imitación, cuando no de parodia. Primero se intentó definir el estilo del autor, aunque sin proporcionar un criterio o una definición para el estilo; parece ser que se definieron algunos rasgos elementales del discurso y técnicas narrativas recurrentes (Garrido-Merchán et al., “*Simulating horror literature by H.P. Lovecraft with ChatGPT*” 6). Todos los hallazgos “estilísticos” fueron comunicados a ChatGPT y se le dejó en disposición de crear historias imitando el estilo del autor. El ejercicio parece bastante sencillo, pero dentro de las conclusiones del artículo se enuncia que para los textos producidos haría falta la revisión final de un experto en Lovecraft, con la finalidad de verificar posibles errores o incoherencias en el proceso y el resultado: “La principal limitante de este estudio es que requiere una muy directa y especializada intervención humana”³⁶ (10).

Pero es *Los campos electromagnéticos* el estudio de caso más interesante. El libro es amplio, si bien no extenso, y de diferentes matices. Como hicimos con el fenómeno de The Policeman’s Beard..., recuperaremos aquí únicamente lo que ha dicho Carrión en la introducción.

La intervención cultural que propone *Los campos electromagnéticos* también es colectiva, también es una invocación, también da testimonio de una liberación brutal y de un giro histórico. Pero ya no se trata de transformar en literatura una dimensión propia de la psique;

³⁵ Original: *Our aim is to create a stimulating environment that fosters co-creation: ideally, the machine should output valuable suggestions, to which the author retains a significant stake within the creative process.*

³⁶ Original: *The main limitation of this study is that it requires very direct and specialized human intervention.*



sino de invitar definitivamente a nuestros exocerebros, a nuestros aliados tecnológicos, a nuestras inteligencias artificiales y compañeras a participar del viejo arte de contar historias y desarrollar ideas y construir belleza, para que escriban con nosotros o incluso más allá de nosotros, como máquinas de escribir autónomas, como procesadores de texto que teclean solos, con o sin nuestra ayuda (Carrión et al. 19).

Más allá de esto, la introducción reconoce las habilidades de las IA y cómo estas pueden actuar en beneficio para los escritores humanos. Es decir, lo que tiene que ver con análisis y el arrojo de datos, por ejemplo. Pero Carrión también está al tanto de las críticas que se le pueden hacer a estas tecnologías en cuanto a los déficits creativos, y desde aquí va estirando una disertación que intenta hacer a un lado esta y otras posibles críticas, en favor del uso de estas tecnologías, en una defensa que parecería ser demasiado entusiasta cuando no ingenua:

La crítica evidente al GPT-3 es que carece de conciencia. Se trata de producción industrial de lenguaje. No hay intención ni comprensión en los enunciados que van surgiendo en la interfaz de *OpenAI*. Pero intuyo que no importa la calidad del sentido original, sino la cantidad de información que ha sido necesaria para crear la posibilidad de que un lector — humano o no humano — encuentre su propio sentido (36).

De haberse ahondado en este problema, el libro figuraría en el primer apartado de este estado del arte. Pero no hace más que replicar a sus otros argumentos, erigiendo esta posibilidad de que un lector pueda dar sentido a lo escrito por una IA a manera de un revés a las críticas. Proporcionar aquí esta libertad, por no decir carga, al lector no implica que se asegure el fallo por la aceptación de esta literatura. Es una cuestión meramente de libertad interpretativa.

Tras su larga introducción, pues, se puede encontrar un libro donde ChatGPT, los modelos 2 y 3, toman la palabra creativa, en un intento similar a lo que se hizo con RACTER. El proyecto



final ahora ser un cambio total en la literatura, en el quehacer literario, en el control mismo de la literatura, que el humano debería ir soltando paulatinamente. O tal vez, según lo sugiere Carrión, aprenda a convivir en armonía con un nuevo tipo de literatura, o de autor, quizá:

Con la ayuda de otras máquinas, fiscalizamos los contenidos que se producen con inteligencia artificial y que simulan realidades alternativas, como las noticias falsas y las falsificaciones profundas. Y corregimos sus textos. Se invierten los roles tradicionales. Si durante un siglo la máquina de escribir, los procesadores de texto o los correctores informáticos han sido nuestros asistentes, ahora nosotros somos los asistentes de la máquina ... Solo así, estudiando ... las redes neuronales de la inteligencia artificial ... para transformarlas en ficciones y en poesía, podremos renovar la narrativa humana en pleno Antropoceno. Por eso aquí, en este volumen colectivo, les cedo la palabra (41-42).

Como este, hay otros varios ejercicios que conjuntan las inteligencias artificiales con la creación literaria, en un intento de consolidar ambas disciplinas. ¿Sería acaso esta la mejor postura a adoptar para mantener a la literatura a la vanguardia? Carrión, por ejemplo, no parece ver la necesidad de esta vanguardia, pero nada nos impide leer que de hecho pueda ser una posibilidad. Por ejemplo, traigamos a colación el ejercicio de instalación o performance en la plaza Trafalgar, Londres, en 2018. El ejercicio consistía en permitir a los transeúntes el acercarse a una interfaz contenida en las fauces de figuras de leones colocadas en esta plaza, e interactuar ingresando palabras con las cuales la IA integrada generaría poemas nuevos, pero con el estilo del Bardo de Avon. La autora del artículo que rescata este ejemplo menciona al respecto que no hace falta entrar en la discusión de una IA que sustituye al ser humano, cuando se pueden obtener ventajas técnicas y aun lúdicas de estas. "...las humanidades tienen abundante material para nutrir estas súper



mentes³⁷ y aprender en ellas (Bucio, “Shakespeare y los leones: encuentro entre inteligencia artificial y humanidades”).

Recuperar estas visiones para el estado del arte nos lleva a asimilar las formas de pensar en la colaboración de las IAG con la literatura. Y si bien las visiones que se han expuesto no son compartidas con las de esta tesis, puesto que no van acompañadas de un criterio más amplio o problematizaciones más profundas, no hay razones específicas por las cuales demeritarlas.

Ahora bien, así como existen estas propuestas optimistas, también hay un lado crítico de esta cuestión, y es el que analizaremos a continuación, de la mano de la autora Lea Henrickson.

La autora construye un aparato crítico sobre la base de la hermenéutica, o lo que ella llama el “contrato hermenéutico”: “En lugar de intentar enumerar las innumerables interpretaciones sobre la autoría, aquí nos enfocaremos en la autoría en lo que respecta a la relación entre el autor y el lector (el contrato hermenéutico)”³⁸ (“*Autorship in computer-generated texts*” 4). Esta postura es interesante en tanto que el método de la autora es no preocuparse tanto por el concepto de autoría sino por ese diálogo que se establece entre autor (nosotros diríamos texto) y lector. Del mismo modo, parece ser que la autora aboga por ver el exterior del texto, y así, en su artículo “*Constructing the Other Half of The Policeman’s Beard*”, argumenta sobre la existencia de factores extraliterarios que vuelven atractivas las lecturas que se puedan hacer de los materiales de las IA: “...la importancia de *The Policeman’s Beard* no radica tanto en sus contribuciones literarias o tecnológicas o literarias, sino en lo que representa en contextos culturales más amplios”³⁹ (11). Esto juega un paralelo relevante con las teorías de Umberto Eco para la interpretación de un texto,

³⁷ Es decir, las IA.

³⁸ Original: *Rather than attempt to list the countless interpretations of authorship, focus here is on authorship as it pertains to the relationship between author and reader (the hermeneutic contract)*.

³⁹ Original: ...made algorithmically authored texts accessible to a general public who, until the book’s publication, had little access to such material. For this reason, *The Policeman’s Beard’s importance lies not so much in its literary or technological or literary contributions, but in what it represents in wider cultural contexts*.



basadas en las estrategias del Autor Modelo y el Lector Modelo, pero la metodología no es semiótica, y solo daremos cuenta de esta metodología de Henrickson. Más nos interesa la función que intenta cumplir con ella. La visión del contrato hermenéutico le ha permitido a la autora alcanzar conclusiones como esta:

A pesar de la mejora tanto de la velocidad como de la potencia de NLG⁴⁰, los lectores no son necesariamente tan receptivos a este nuevo medio de producción de texto. De hecho, la importancia cultural de NLG no parece estar tan relacionada con mejorar la capacidad de personalización textual, sino más bien con resaltar el deseo de los lectores de una conexión entre humanos. Los NLG cuestionan la comprensión convencional de la autoría, con cierta sensación de que esta tecnología es inapropiada para las circunstancias culturales actuales, ofendidos por la amenaza potencial que representa la comunicación entre humanos y computadoras para la conexión interpersonal. Por supuesto, es probable que estas actitudes no permanezcan estancadas. Sin embargo, en este momento los lectores buscan lo humano en textos generados por computadora, buscando una conexión con un autor imaginado para tener una sensación de experiencia compartida. Quién —o qué— es ese autor sigue siendo objeto de debate⁴¹ (“*Autorship in computer-generated texts*” 13).

Aun y con estas observaciones, en un artículo posterior la autora parece querer abrirse a la idea de las máquinas creativas. En “*Prompting meaning: a hermeneutic approach to optimising*

⁴⁰ NLG: modelos de *Natural Language Generation*. Lo mismo que con los LLM: *Large Language Models*, consideramos estos modelos en general como inteligencias artificiales generativas.

⁴¹ Original: *Despite NLG's enhancement of both speed and power, though readers are not necessarily so receptive to this new means of text production. Indeed, the cultural significance of NLG seems not so related to enhancing capacity for textual personalization, but more to highlighting readers' desire for human-to-human connection. NLG brings conventional understandings of authorship into question, with some feeling that this technology is inappropriate for current cultural circumstances, offended by the potential threat posed to interpersonal connection by human-computer communication. These attitudes, of course, will likely not remain stagnant. Right now, though, readers search for the human in computer-generated texts, seeking a connection with an imagined author for a sense of shared experience. Just who – or what – that author is remains up for debate.*



prompt engineering with ChatGPT", continúa apostando por el contrato hermenéutico, pero sus argumentos se muestran más flexibles cuando admite que el lector podría, a fin de cuentas, terminar de aceptar las creaciones de las IA al punto de llegar a disfrutarlas.

...hemos demostrado que los lectores siempre encuentran significado a partir del texto, independientemente de si ese texto es producido por un ser humano o por una IA. Cuando los lectores pueden solicitar la generación de textos ellos mismos, pueden utilizar la ingeniería de prompts para lograr resultados que sean significativos para ellos de alguna manera, ya sea haciéndolos reír, inspirándoles nuevas ideas, ayudándolos a comprender un tema o cumpliendo cualquier otro propósito. El significado proviene de lo que los lectores hagan con las respuestas a sus preguntas⁴² (14).

Como sucedió antes con Carrión, el señalamiento no es erróneo, sino solo incompleto: al decir que el lector tiene la capacidad —y aun la tarea— de completar el significado de una obra, hay cuestiones mucho más complicadas que esto, que Henrickson no aborda o no prevé. Solo para darnos una idea de lo problemático de la cuestión, volvamos a Ertel. El siguiente es un postulado que, de nuevo, intenta argumentar sobre la posibilidad de la creatividad de las IA, pero exhibiendo un problema fundamental:

Imaginemos un programa que genera texto aleatoriamente letra por letra. El resultado podría comenzar así:

NitDjgMQsQfI 6zz1B:6xZkgp1xe.EqyD7z(C.

⁴² Original: ...we have shown that readers are always making meaning from text, regardless of whether that text is produced by a human or AI. When readers are able to request the generation of texts themselves, they may use prompt engineering to achieve output that is meaningful to them in some way, whether by making them laugh, inspiring new ideas, helping them understand a subject, or fulfilling any other purpose. Meaning comes from whatever readers will make out of the answers to their prompts.



Si bien el texto es nuevo, no es muy valioso y, por lo tanto, no es un artefacto creativo. Por otro lado, un programa que consulta y genera texto desde una base de datos tampoco cuenta como creativo porque no produce nada nuevo⁴³. Un programa creativo, por ejemplo, debería escribir una novela que sea interesante y nueva al mismo tiempo ... Lo interesante es más difícil. Sería bueno si un programa pudiera generar nuevos textos que una persona disfrutaría basándose en las calificaciones que esa persona ha otorgado a la literatura leída anteriormente⁴⁴ (282).

Esto conecta con la idea de Henrickson en el sentido de que es posible tomar en cuenta la consideración del lector para definir lo que resulta interesante y establecer parámetros. Pero al centrarse únicamente en buscar una respuesta que involucre al lector, el problema del texto creativo y no creativo que plantea Ertel, o la visión de los textos de las IA como textos pragmáticos que son aún más interesantes en lo extraliterario, no queda satisfecha.

Similar a lo que expusimos en el planteamiento del problema, dentro de este estado del arte crítico consideramos abordar también casos donde los receptores de los textos de las IA expresan cierta inconformidad ante ellos. Por ejemplo, uno de los desarrolladores de una IA llamada “NeuroWriter”, programada para imitar los estilos literarios de varios autores y escribir guiones de cine, dijo lo siguiente para una entrevista: “Nuestra intención no es publicar un *bestseller* ni hacerles competencia a los escritores reales” (Barbero, 2018). Su objetivo es más bien evaluar cómo trabajan los algoritmos los datos entregados y continuar haciendo experimentos con guiones,

⁴³ Ertel parecería caer aquí en una contradicción. En la cita que recuperamos para el planteamiento del problema, el autor menciona que un programa puede producir texto desde cero (*from scratch*) aun desde una base de datos.

⁴⁴ Original: *Let us imagine a program that randomly generates text letter by letter. The result might begin something like // NitDjgMQsQfI 6zzIB:6xZkgpIx.e.EqyD7z(C. // While the text is new, it is not very valuable and therefore not a creative artifact. On the other hand, a program that queries and outputs text from a database does not count as creative either because it produces nothing new. A creative program, for example, should write a novel that is both interesting and new at the same time. (...) Interestingness is more difficult. It would be good if a program could generate new texts that a person would enjoy based on the ratings which that person has given to previously read literature.*



a los que en algún punto se refirió como “delirantes”. De confirmarse que no hay intenciones de interferencia, el ejercicio permanece como eso, un mero ejercicio técnico, y habría entonces que separar estas cuestiones de aquellas donde las creaciones de las IAG se lanzan a los mercados.

Un artículo de *The Guardian*, escrito por Rachel David, recoge las opiniones de dos artistas y dos científicos computacionales. Tomamos como muestra las declaraciones de dos de ellos, uno de cada disciplina. Erica Wagner, autora y crítica literaria, dijo: “Creo que nuestros cerebros humanos son mucho más sutiles, extraordinarios y extraños de lo que las máquinas podrían ser. Yo diría eso, ¿no? No quiero que me dejen sin trabajo. La mayoría de la gente pensaba que nadie jamás volaría en una máquina, por lo que debemos tener mucho cuidado al decir de qué creemos que son capaces las máquinas.”⁴⁵ Por su parte, Michael Osborne, catedrático de la materia de “machine learning” (aprendizaje de máquinas) en Oxford, se mostró menos preocupado: “La creatividad sin dudas es la facultad humana más difícil de automatizar: es poco probable que los robots sean completamente creativos en el corto plazo ... Ciertamente es posible diseñar un algoritmo que pueda producir una secuencia interminable de pinturas, pero es difícil enseñarle a dicho algoritmo la diferencia entre lo emocionalmente poderoso y lo malo ... Aportando algo de evidencia a nuestra tesis, encontramos, tanto para el Reino Unido como para los EE.UU., que casi el 90% de los empleos creativos tienen un riesgo bajo o nulo de automatización.”⁴⁶.

Henrickson con su visión captura una parte importante del problema, pero sería de esperarse que un aparato crítico decidido a no prestar más atención al problema de la autoría para

⁴⁵ Original: *I think that our human brains are much more subtle and extraordinary and strange than machines could ever be. I would say that, wouldn't I? I don't want to be put out of a job. Most people thought that no one would ever fly in a machine, so we have to be pretty careful about saying what we think machines are capable of.*

⁴⁶ Original: *Creativity is arguably the most difficult human faculty to automate: robots are unlikely to be fully creative any time soon. (...) It is certainly possible to design an algorithm that can churn out an endless sequence of paintings, but it is difficult to teach such an algorithm the difference between the emotionally powerful and the dreck. (...) Putting some evidence to our thesis, we found, for both the UK or the US, that almost 90% of creative jobs are at low or no risk of automation*



centrarse en el del lector, es decir, una cuestión hermenéutica, profundice en su labor de abordar una cantidad importante de fenómenos, y que termine de definir la calidad de ese contrato hermenéutico entre máquina y humano. Quizá en trabajos posteriores Henrickson pueda encargarse de ello. Entre tanto, las problematizaciones siguen abiertas. Asimismo, las visiones optimistas de Carrión y los otros autores cuentan con una visión tal vez más sólida respaldada por su optimismo.

Quizá sea esta ambivalencia en la visión de las IAG desde la literatura la que luego lleva a la generación de estudios para evaluar la relevancia y el atractivo de estas obras. En el planteamiento del problema hablamos de un par de estudios aplicados a poesía, pero el siguiente, de Vivian Emily Gunser y otros, titulado “*The Pure Poet: How Good is the Subjective Credibility and Stylistic Quality of Literary Short Texts Written with an Artificial Intelligence Tool as Compared to Texts Written by Human Authors?*”, enfocado a cuentos (*short stories*) puede ser incluso más revelador para el caso de la tesis. Lleva a cabo un doble estudio y corrobora algo ya establecido: a) la preferencia por los cuentos de seres humanos antes que los de las IA. En general, los cuentos fueron señalados por profesar una mejor calidad estética y estilística (1749); y b) que las IA podrían convertirse en herramientas para la escritura creativa, colaborando con la parte argumentativa, por ejemplo (1749). (Es por esta puntualidad en lo literario que hemos separado este estudio de los otros que llegaban a estas mismas conclusiones). Aun así, Gunser y compañía, no pierden de vista que incluso esto tiene sus limitantes:

Sin embargo, al considerar las herramientas de IA como socios de escritura colaborativa, es importante que el uso de la herramienta de IA constituya una mejora de la creatividad en lugar de un reemplazo de los humanos en el proceso de escritura (cf. IA centrada en el ser humano; Shneiderman, 2020). También es importante tener en cuenta que todavía no



existen directrices legales que puedan considerarse autores de este tipo de obras literarias generadas por IA⁴⁷ (1749).

Este estudio que parte más bien desde la literatura nos interesa por estar bastante “completo”: lleva a cabo un estudio y concluye algo a partir de este no sólo en un intento de ver las IA como herramientas, pero también desde lo ético. Es una visión más o menos compartida con los objetivos de esta tesis. Su agregado al estado del arte representa otro añadido crítico que acompaña a las conclusiones de Carrión y Henrickson especialmente.

Aun con todo esto, quizá la visión que mejor pueda rescatar este estado del arte es la de los estudios más próximos a la narratología. Lo que en estos se muestra como más incisivo es, por obvias razones, el señalamiento de los puntos débiles, por llamarlos de algún modo, de las IA, pero bajo una terminología más ajustada, hecho que consideramos útil.

El grupo de artículos “Literatura, narratividad y composición en la era de la Inteligencia Artificial”, “Narrative Models: Narratology Meets Artificial Intelligence” y “The Art of Narration and Artificial Narrative Intelligence: Implications for Interdisciplinary Research”, comparten el objetivo de describir algunas de las cuestiones de por qué las inteligencias artificiales generativas enfrentan visibles dificultades para producir historias bajo el criterio convencional de “interesantes” o “de calidad”. Ya sea por la falta de experiencia y subjetividad (Arathdar, Debarshi 381), que deja en desventaja a las IA frente a los seres humanos, resaltando su cualidad “mecánica”, o por la falta de competencias pragmáticas —y semánticas— en general (Adamivna 314), que delatan una visión más bien lógica y matemática de la literatura (315). Estas dos autorías

⁴⁷ Original: *However, when considering the AI-tools as collaborative writing partners it is important that the use of the AI-tool should constitute an enhancement in creativity rather than a replacement of the humans in the writing process (cf. human-centered AI; Shneiderman, 2020). It is important to also keep in mind that there are still no legal guidelines who can call themselves the author of such AI-generated literary works (Gunser et al. “The Pure Poet ... ”) (1749).*



esbozan brevemente propuestas quizá demasiado ambiciosas. Lo mismo se lleva a cabo en el estudio de Gervás et al.: “*Narrative Models: Narratology Meets Artificial Intelligence*”. De este citamos este esbozo de propuesta por representar esa visión compartida con las otras autorías:

En nuestra perspectiva, una combinación metodológica de descripción, análisis y generación —en otras palabras: un enfoque interdisciplinario— tiene el potencial de lograr un avance cualitativo mutuamente beneficioso en la investigación sobre la generación de historias y sobre los modelos narrativos en general. Este enfoque interdisciplinario podría comenzar identificando aquellos modelos narrativos existentes en las humanidades cuyo conjunto de compromisos ontológicos se adapta mejor a la tarea de generación de historias, y buscando (o produciendo) implementaciones de estos modelos orientados computacionalmente⁴⁸ (7).

De los tres artículos, el de Gervás se muestra más elocuente en la labor que podría llevarse a cabo, una labor, definitivamente, interdisciplinar. Y no obstante que la propuesta quede también como demasiado prometedora, el autor es también más incisivo en abordar cómo al menos ciertos tipos de IAG utilizan modelos narrativos y describe un poco de estos, yendo más allá que las propuestas de los documentos del primer apartado: “La inteligencia artificial utiliza dos técnicas en la generación de historias: planificación/resolución de problemas y gramáticas de producción. Las reglas específicas utilizadas en sus algoritmos podrían verse influenciadas por conocimientos de estudios literarios u otros campos (por ejemplo, psicología de la lectura y la escritura)”⁴⁹ (1).

⁴⁸ Original: *In our view, a methodological combination of description, analysis and generation – in other words: an inter-disciplinary approach – holds the potential for a mutually beneficial qualitative breakthrough in research on Story Generation, and on narrative models in general. This interdisciplinary approach might start by identifying those existing narrative models in the Humanities whose set of ontological commitments is better suited for the Story Generation task, and by searching for (or producing) computationally oriented implementations of these models (Gervás et al., 7).*

⁴⁹ Original: *Artificial Intelligence uses two techniques in Story Generation: planning/problem solving, and production grammars. Specific rules used in their algorithms might be influenced by insights from literary studies or other fields (e.g. psychology of reading and writing).*



Páginas antes menciona que estos modelos estarían basados en la morfología de Propp (1), y que en todo caso estos modelos podrían valerse de otras teorías literarias (el artículo trae a colación a Todorov, Genette y van Dijk, páginas 3-4). Finalmente, como la mayoría de estudios, encuentra el fin del camino, es decir, que los modelos narrativos implícitos en las IAG estudiadas “...se basan generalmente en un concepto altamente reduccionista de “historia” que ignora los conocimientos de las disciplinas de humanidades sobre la complejidad y la dinámica de la narrativa”⁵⁰ (7),, el cual le lleva a formular su propuesta conclusiva que hemos citado antes.

Todavía podemos aludir a una última conclusión, encontrada por Marie-Lauren Ryan en su libro, *Possible Worlds, Artificial Intelligence, and Narrative Theory*, obra cuyo postulado principal es un inverso de lo que hemos venido examinando: no se trata de cómo aporta la teoría narrativa a las IA, sino lo contrario: cómo estas aportan a la comprensión de la construcción de mundos posibles y otros artificios narrativos. Con este caso nos hemos tomado una “licencia” en cuanto a la delimitación de años establecida en un principio, ya que, como se dijo, existen estudios más remotos directamente relacionados con lo que se pretende en esta investigación documental. Sin embargo, para no contaminar las observaciones precedentes, rescataremos de este libro una visión general sobre la IA, pertinente y válida todavía para los modelos de IAG más avanzados de tiempos actuales.

Se trata de un criterio tanto estético como semántico que toma de Paul Ricoeur, y es la *falta de inteligencia narrativa* que se verifica en las IAG: “... el hecho de que algo que se asimile a la trama de una historia bien conocida vino de una computadora, no es prueba de que la máquina en

⁵⁰ Original: *Artificial Intelligence approaches in Story Generation are generally based on a highly reductionist concept of ‘story’ which ignores the Humanities’ disciplines insights into the complexity and dynamics of narrative (...)*



verdad exhiba lo que Paul Ricoeur (1982) aptamente llamó ‘inteligencia narrativa’⁵¹” (Ryan, 233).

Pero lo más interesante de este criterio, que funciona como una suerte de “evaluación” para las creaciones de las IAG, es cómo ejemplifica esto:

Si un programa recibe como entrada el léxico de la lengua española, eventualmente produciría el *Quijote* simplemente por intentar todas las combinaciones de palabras posibles. Pero este programa no tendría entendimiento alguno de la trama que genera, y sería incapaz de separar de su producción masiva los intentos fallidos de aquellos que resultaron en historias viables. Lo que le faltaría a este programa es un método heurístico “capaz de dirigir la búsqueda (de la historia) por los caminos con más probabilidades de éxito” (Boden 1977: 347). De manera alternativa, uno podría imaginar un programa que no recibiría ningún aporte en absoluto, pero que consistiría en procedimientos tan específicos que todas sus ejecuciones producirían inevitablemente el texto de *Don Quijote*, junto con una representación semántica explícita de los eventos narrativos. Este programa lograría una perfecta comprensión de su resultado (...)⁵² (233).

Ryan no solo está aportando su “pequeño grano de arena” a la discusión de las otras autorías —y de hecho es revelador que aunque el libro fue escrito más de tres décadas atrás, las propuestas siguen apareciendo como necesarias, dando a entender que no se han atendido idealmente—, sino que está esbozando un problema fundamental que será comentado ampliamente en su momento en

⁵¹ Original: *the fact that something resembling the plot of a well-known fable came out of a computer is no proof that the machine actually displayed what Paul Ricoeur (1982) aptly calls "narrative intelligence".*

⁵² Original: *If a program received as input the lexicon of the Spanish language, it would eventually come up with the text of Don Quixote by simply trying out all the possible word combinations. But this program would have no understanding of the plots it generates, and it would be unable to weed out, from its massive output, the unsuccessful tries from those which resulted in viable stories. What this program would lack is a heuristic method "able to direct the search (for the story) along the paths most likely to succeed" (Boden 1977: 347). Alternatively, one could imagine a program that would receive no input at all, but would consist of such specific procedures that all of its runs would inevitably produce the text of Don Quixote, together with an explicit semantic representation of the narrative events. This program would achieve a perfect understanding of its output (...)*



esta tesis: las IA pueden tener muy buenas competencias sintácticas, mas no pragmáticas ni semánticas. Y es precisamente este argumento, planteado desde la literatura, el que mayor fuerza puede cobrar, el que bordea a la mayoría de las demás observaciones, como las de los modelos narrativos. En suma, el hallazgo de Ryan vuelve a la piedra angular del problema narrativo-literario de las IA, el “talón de Aquiles” de estas.

Los artículos y libros aquí expuestos son ejemplares en términos de los encuentros de áreas más especializadas de la teoría literaria y la inteligencia artificial. A través de estos observamos lo que se está hablando cuando se unen estos campos específicos: se analizan las creaciones de las máquinas, sus herramientas, y se llega al fallo de que aún hay una gran distancia que separa las posibilidades creativas de las IAG de las de los seres humanos, comparativa que se ejerce en la mayoría de artículo. Por el valor y calidad metodológica de algunos de estos artículos, volveremos a ellos más adelante.

4. La ética del uso de inteligencias artificiales para las creaciones literarias

La cuarta y última categoría de este E.A.C va sobre la ética, en tanto que hemos propuesto en el planteamiento del problema que la mirada de la ética para nuestro objetivo es indispensable, como lo es para el tema en general de las IA en las Humanidades.

En una recopilación documental a fondo que intente unificar las categorías de literatura, ética e inteligencia artificial, lo que emerge es una discusión más o menos exhaustiva sobre la autoría y el plagio. “Más o menos” en tanto que ya hemos examinado en los apartados anteriores estudios donde eventualmente se llega a un bloqueo en los avances, donde el estado del arte parece haber encontrado sus límites respecto a lo que han hecho ya las autorías. Parece haber, de manera general, dos consensos: uno, que los autores de las creaciones de las IA son siempre los agentes



humanos involucrados ora en la programación o creación de las IA; dos, que pese a esto la cuestión todavía no se ha resuelto y no hay fallos definitivos, al menos hasta el momento de la presente redacción.

Los artículos “*Ethical challenges of artificially intelligent poetic works: subjectivity, ownership and cultural transformation*”, “La creatividad computacional y la protección autoral de las obras generadas por inteligencia artificial”, “¿Autoría algorítmica? Consideraciones sobre la autoría de las obras generadas por inteligencia artificial y “Autoría e inteligencia artificial generativa: presupuestos filosóficos de la función del autor”, hacen de sus puntos centrales la autoría y la propiedad intelectual de las IA, aportando visiones distintas cada uno por lo cual debería de rechazarse la aplicación de este concepto. El primer artículo establece como su razón principal las faltas de subjetividad y emociones de la IA (Bin, Hu 5002), y llega a la conclusión de que la autoría debería otorgarse a los “participantes humanos como desarrolladores, proveedores de datos de capacitación o usuarios” (5003). Esta manera de procesar la cuestión es interesante, pero por sí solos esos factores no terminan de concatenar, en el artículo de Hu Bin, con la idea de por qué una autoría conlleva esas nociones para poder concretarse. Y ciertamente no ha sido de todo un impedimento para llevar a cabo *simulaciones* en las que máquinas obtienen estas autorías, del mismo modo que la mencionada antes robot Sofía obtuvo una ciudadanía. Desde un punto de vista filosófico general, la discusión habría resultado más fructífera si el autor hubiera enlazado sus motivos principales con la mención que también hace respecto a que las IA, con sus algoritmos, se basan en la imitación de obras existentes (4994).

El investigador Luis Vázquez ha sido el encargado de los dos artículos siguientes en la lista. A diferencia de Bin, propone que los criterios para la autoría son, básicamente, el ser humano y poseer las cualidades de intelecto (101). Esta noción es problemática también filosóficamente,



pero no ahondaremos en ella. De este artículo solo rescatamos otro señalamiento de valor, el de que actualmente las obras que se producen con IA son parte del dominio público, según la mayoría de las legislaciones en este caso (116). En su segundo artículo refuerza la idea principal, zanjando todavía más que Bin la cuestión al porvenir (231).

En “Autoría e inteligencia artificial generativa…”, el criterio para otorgar una autoría es la función social del autor, lo cual también conlleva implicaciones sobre la toma de decisiones conscientes y lo que se señala como “responsabilidad estética”:

A un autor le atribuimos decisiones conscientes de las que derivamos su responsabilidad jurídica, pero también su responsabilidad estética con la que identificamos un estilo. Por ello un autor es más que una firma o un nombre. Siguiendo a Foucault, se afirmó que la función de autor es una forma de despliegue de la subjetividad que, además de implicar la perspectiva de primera persona, implica por ello un nombre, una relación de apropiación y de atribución y un marco de referencia para el lector (Misseri 250).

La base teórica de Misseri es decididamente filosófica, con esa tracción desde Foucault, y si bien la discusión obviamente no cierra aquí, su fundamentación parece más sólida que las anteriores. Al menos aquí la subjetividad está en función de demostrar algo más, y no se toma como un concepto aislado entre otros.

Es evidente que este grupo de artículos tiene por objetivo cuestionar la noción de la autoría y propiedad intelectual en las inteligencias artificiales. Por ello proponen un marco teórico desde el cual emitir sus argumentos. Por otro lado, sin embargo, existen pensamientos con otro tipo de propuestas. A saber: la de permitir un tipo de autoría especial a las IA. El artículo titulado “*Writing at a Distance: Notes on Authorship and Artificial Intelligence*”, comparte una visión similar a otras expuestas donde los agentes humanos detrás de las IA podrían adjudicarse las creaciones de estas,



con la gran diferencia de que propone un tipo de autoría que define como “autoría distribuida” (*distributed authorship*). Dicho proceder, parte de una necesidad de actualización o resignificación de ciertos conceptos, como resultado de la corriente que emite una época en la que las máquinas profesan la capacidad de imitar al ser humano (Bajohr, Hannes).

Pero todavía queda examinar propuestas que recuperen y/o creen marcos éticos para tratar cuestiones de las IA más generales. Para este aspecto recuperamos una única propuesta, dado que este examen requeriría la salida de los límites que establecimos con las categorías de esta investigación, pero hemos decidido incluirla por la razón de que un análisis desde la literatura podría retraeer esta propuesta desde los límites hacia el centro.

La propuesta es la de Ben Shneiderman y su libro de reciente publicación (2022) titulado Human-Centered AI. El título viene dado por una rama de estudio de la IA que, como su nombre lo dice, toma en cuenta al ser humano en tanto que este debería priorizarse en su accionar, bienestar y libertad a la actuación de las IA (7). Lo que de momento requerimos para el estado del arte es reconocer la metodología que ha servido al autor para hacer valer sus argumentos. Se trata de las estructuras de gobernanza. El autor observa que estas estructuras permitirían reducir las brechas entre las discusiones éticas y los pasos para su aplicación práctica en lo que respecta a las inteligencias artificiales. Emite, pues, tres señalamientos, de entre los varios que se han discutido en estas estructuras de gobernanza, de los que estarían a cargo tres principales agentes, cuyas decisiones tienen un alto impacto en lo que se puede lograr hacer con las inteligencias artificiales: líderes de equipos de software, gerentes de negocios y líderes de varios tipos de organizaciones.

Los líderes de equipos de software, gerentes comerciales y líderes de organizaciones tendrán que adaptar prácticas técnicas comprobadas, estrategias de gestión y métodos de supervisión independientes, para poder alcanzar los objetivos deseados de:



- 1) Sistemas confiables basados en prácticas comprobadas de ingeniería de software;
- 2) cultura de seguridad a través de estrategias de gestión empresarial; y
- 3) certificación confiable mediante supervisión independiente y regulación gubernamental⁵³ (10).

La idea es que al llevar a cabo la aplicación de esta visión de las HCAI, se dará paso a una mejor convivencia entre los humanos y las tecnologías con inteligencia artificial. La creatividad es, por supuesto, mencionada en la ecuación, aun si de manera general: “Los investigadores y desarrolladores de sistemas HCAI 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”⁵⁴ (7). Es por eso que la cuestión se abre a ser planteada desde y para la literatura. Pero la visión de las HCAI no es precisamente mantener a raya la participación de estas tecnologías en la vida de los seres humanos: todo lo contrario. Se sugieren altos niveles de automatización —para ciertas áreas— siempre y cuando el ser humano tenga el control de esta alta automatización. Así, pues, la aspiración final de las HCAI es enlazar esta dicotomía, “tener altos niveles de control humano y altos niveles de automatización”⁵⁵ (9; ver también en 59-60). Esta idea será explorada a detalle en el capítulo tres de esta tesis.

Esto es lo que se ha discutido en la visión general de la ética de las IAG en la creación literaria: imperan las discusiones sobre la autoría de los textos generados con IA, y esto nos lleva

⁵³ Original: *Software team leaders, business managers, and organization leaders will have to adapt proven technical practices, management strategies, and independent oversight methods, so they can achieve the desired goals of: // 1) Reliable systems based on proven software engineering practices; // 2) Safety culture through business management strategies; and // 3) Trustworthy certification by independent oversight and government regulation.*

⁵⁴ Original: *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.*

⁵⁵ Original: *The new aspiration is to have high levels of human control AND high levels of automation.*



también a repensar el papel no sólo ya de los autores, sino de los seres humanos en su relación general con la tecnología. Por otro lado, estas propuestas también son útiles al momento de establecer críticas relevantes sobre las IA en cuanto a las problematizaciones establecidas en la Figura 1, en especial en lo que a la influencia en la toma de decisiones de los seres humanos respecta. Y es en este punto donde la literatura hace su entrada en muchas maneras. Una de ellas, de hecho, la que problematiza esta tesis, y que se centra en el lector. Por eso habremos de volver a pensar en la ética más adelante.

Conclusiones

El estado del arte crítico ha sido de utilidad para trazar un panorama basto sobre los diferentes conceptos, argumentos, metodologías y particularidades que acompañan el tema de las inteligencias artificiales en su encuentro con la literatura. A manera de síntesis, delinearemos las conclusiones más relevantes.

La primera anotación fue respecto a la abundancia del material, hecho que volvió a la selección y delimitación de estos desafiantes. En cuanto a los encuentros entre la teoría de Eco, su Autor Modelo y Lector Modelo, como estrategias para el encuentro del lector con un texto, y las IAG, lo que más pudimos observar fueron modelos semióticos de aplicación, no un tratamiento directo de esta teoría. Esto nos ha llevado a concluir que el problema del lector ante el texto de la IAG, con todas sus implicaciones, continúa sin ser explorado adecuadamente.

Tras de esto, obtuvimos el panorama sobre la categoría de la creatividad, observando de primera instancia que esta no profesaba de una definición bien establecida, lo cual afectaba el resto de las discusiones, manteniéndolas en la superficie del problema. Pero es ya de anotar que la



creatividad, como sea que se defina, afectada por las IA, es un tema cuyo interés se revela desde la cantidad de artículos disponibles en la red.

El siguiente grupo, más amplio, demostró el interés que hay por el estudio de las capacidades literarias de las IAG. Destacaron por sobre todo la ausencia de razones justificadas por las cuáles se decidía “nutrir” las capacidades de las IA, y las observaciones de Henrickson sobre la autoría y la metodología de la hermenéutica como herramienta principal para la mediación lector-texto producido con IAG, una metodología que, por otro lado, no colaboró demasiado en la profundización del tema. Se observó también que las capacidades literarias de las IAG se encuentran todavía muy lejos de alcanzar un punto satisfactorio o interesante siquiera, pero se dio a entender que se prevén maneras de lograr dar un salto hacia este objetivo, involucrando, ambiguamente, más de las áreas de las Humanidades y la literatura. Las razones de esto siguen quedando en el aire, pero nos inclinamos a deducir que el mayor beneficio lo obtienen las disciplinas informáticas y de la IA.

Finalmente, los textos encontrados con la categoría de la ética nos llevaron por la ruta que se pregunta por los derechos autoriales y la figura del autor empírico. No obstante la pertinencia del tema, llama la atención que no se encontró un documento que profundizara o problematizara una ética en las IAG para el uso creativo. Las menciones a esta problemática las podemos encontrar a lo largo de los cuatro apartados, y en libros como el de Shneiderman encontramos una apertura para explorar la cuestión. Pero los problemas que siguen imperando al conjuntar la visión ética de las IAG en la literatura son el de los derechos de autor y la propiedad intelectual, todos ellos como derivados del hueco que deja el plagio.

Estas conclusiones conforman el estado del arte de las inteligencias artificiales en la creación literaria. Nos sirven, entre otras cuestiones, para reconocer los huecos teóricos del objeto



de estudio. Este reconocimiento nos delinea todavía más el camino a seguir para esta tesis. Son tres los huecos principales: el primero, la falta de atención a la figura del lector en la relación con el “texto artificial” de las IAG, el segundo, la falta de un examen estrictamente literario del trabajo creativo de las IAG, el tercer, preguntas y argumentos éticos más incisivos en cuanto a lo que implicaría una dinámica con el lector de los textos de las IAG.



Marco teórico: La semiótica textual de Umberto Eco y su teoría de la cooperación interpretativa con el texto creativo en *Lector in Fabula*.

Introducción

Tanto el estado del arte crítico como los apartados previos se han valido de una parte de las categorías que en este marco teórico se busca expandir, para poder sustentar la teoría que guía a esta tesis y sus hallazgos. Como se menciona en los objetivos, las categorías principales son las de *Autor Modelo* y *Lector Modelo* y *cooperación interpretativa con el texto*, todas de la teoría de la semiótica textual bajo la metodología específica y nomenclaturas de Umberto Eco. Con la finalidad de poder hacer un manejo adecuado de los conceptos, y sistematizar su aparición a lo largo del cuerpo de la tesis, este marco teórico expandirá la conceptualización no solo de estos conceptos, sino del principal andamiaje teórico de la semiótica textual, que está enfocada a definir las particularidades de la dialéctica texto-lector.

La mayor parte de la teoría se condensa en el libro de Eco *Lector in fabula*, sin embargo, como se habrá de aclarar en el apartado metodológico, algunas propuestas de otras teorías se utilizarán en tanto que respaldan y esclarecen lo propuesto en este libro en particular del semiólogo. Una exploración de las particularidades de las IAG en cuanto a su relación con el lector, sugerimos, encuentra su mayor estabilidad y aprovechamiento en una semiótica textual, dado que la metodología que esta sigue, la calidad de sus conceptos, sus bases teóricas y las conclusiones a las que logra llegar, la vuelven una metodología competente y provista de un bagaje de categorías y conceptos propios que podemos definir puntualmente y echar mano de ellos en el camino. Caso contrario, quizá, al de una hermenéutica⁵⁶.

⁵⁶ No apelamos aquí por una oposición entre hermenéutica y semiótica. Nos ceñimos a lo mencionado ya sobre la semiótica como óptima en cuanto a su bagaje categórico y metodológico para explorar un texto. Ricoeur reflexiona



Los conceptos principales por analizarse en este marco teórico son los de Autor Modelo, autor empírico, Lector modelo, lector empírico, obra abierta y obra cerrada, siendo estos en su conjunto los que determinarían la calidad de la categoría de *cooperación interpretativa* del texto de una IAG con el lector. Antes de abordarlos por completo, creemos pertinente exponer de manera general las particularidades de la obra de Eco, en favor de facilitar la comprensión de los conceptos en sí, y de terminar de establecer a qué nos referimos con esta relación del lector con su texto y qué esperamos de ella. De manera que primeroaremos un repaso a las intenciones de Eco para ahondar en esta teoría, una teoría de la recepción simple y llana por un lado, una teoría de la lectura de signos por otro, de índole semiótica. Partiremos, pues, desde su propuesta en el libro *Obra abierta*, y cómo esta no se ve concretada sino hasta el *Lector in fabula*, para a continuación empezar a abrir el panorama sobre la participación del lector en el texto y en la literatura.

De la obra abierta al lector in fabula

En *Obra abierta*, de 1962, Eco abordaba con curiosidad el arte contemporáneo en su facultad de interactuar con el espectador, con el receptor. Su metodología en esta etapa es previa al riguroso abordaje semiótico que más tarde caracterizaría sus obras. Para abordar la cuestión, se valió de la teoría de la información, y de observaciones lingüísticas. No obstante los resultados favorables que de esta pretensión desembocaron, la pregunta continuaba, valga la redundancia, *abierta*. ¿Qué permite a una obra de arte contemporánea —aunque también a una obra de arte en general, de “otros tiempos”— abrirse a la interpretación de su lector (lector en el sentido general

sobre los encuentros entre semiótica y hermenéutica, y menciona el interés de una hermenéutica por una semiótica como “método y técnica de análisis que exige la abstracción del texto” (“Hermenéutica y semiótica” 92), pero abandona la idea de esta en cuanto toma un cariz idealista, apelando a la hipóstasis del texto: “*No* a la semiótica cuando se convierte en la ideología del texto en sí” (92). Bajo esta noción misma noción opera nuestro uso de la semiótica como metodología de análisis de esta tesis.



en que una obra es un texto con un lenguaje propio para ser leído)?— O en su defecto, ¿qué es lo que cierra la obra de arte e impone su interpretación unívoca, dañando así la libertad del lector para entrar en el texto e interpretarlo?

Podemos pensar en una pintura. La observación de esta trascurre varios factores alrededor de su composición que determinan cierta información a su espectador, a su lector. La época es uno de esos factores, y de ese modo, por ejemplo, el *Guernica* no sólo es una pintura de 1937, sino que se encuadra en el gran contexto de la Guerra Civil Española. Es bien conocido en el mundo del arte, y en la cultura, que el *Guernica* es una obra que hace una alusión al Bombardeo de Guernica. Este es el tema de la obra. Independientemente de que Picasso haya dado declaraciones y señalamientos, puntuales o generales, de que su obra es una alusión específica a ese evento, un lector que se enfrente de cara al cuadro y cuente con la información previa de que la obra tiene su génesis creativo en un evento oscuro de la Guerra Civil Española, entrará entonces en un proceso de “intercambio” con la obra. Y es aquí donde inicia lo importante: el cómo la obra, es decir, el color que hay en ella, sus formas, el ritmo, los personajes del cuadro, etc., le llevaría al lector a emitir la interpretación “la obra trata del Bombardeo de Guernica” es un *proceso abierto*. La obra es extensa (tanto en tamaño como en mensaje) y si su mensaje no está dado directamente —pero sin tampoco representar una clave criptográfica— el lector entonces debe entrar en ese proceso de comunicación que es una de las funciones del arte: plantear preguntas, establecer inducciones y deducciones, postular teorías, etcétera, para poder llegar a la conclusión de que por X factores, en efecto, tal cuadro es sobre el Bombardeo de Guernica. Queda claro que en tanto más factores externos conozca el espectador de la obra y su autor (como los ya mencionados, entre los que se puede incluir una noción, aunque sea ligera, de la poética de Picasso y su estilo) más familiares le resultarán algunos elementos y más fácilmente, tal vez, hará una lectura habilidosa de la obra y



alcanzará una interpretación interesante. Si el lector del texto no contara con esa información previa, perdería un indicador importante, pero ello no implica que no pueda llegar a la misma conclusión con el examen de la obra. Como habremos de señalar en constantes ocasiones, una obra, en especial una obra abierta que invite a leerla, a interpretarla, otorgará a su lector ciertas “herramientas”, el lenguaje, la *encyclopedia* cargada de claves importantes —no todas, por supuesto— para que este pueda lanzarse a esa labor. De este modo, la obra es un flujo continuo de comunicación, que si bien se cierra y reposa su movimiento cuando se llega a una conclusión sobre ella, pronto vuelve a animarse, permitiendo más interpretaciones de otros lectores, manteniendo siempre su contenido fresco y renovado y abierto a la comunicación. Esta ductilidad, o bien permisividad, esta voluntad de comunicarse con su lector, su espectador, y que además es una clara demostración de respeto a la inteligencia de estos, que tienen las obras de arte, es lo que Eco denomina la *poética del arte contemporáneo* (*Obra abierta* 73-75, 105).

Caso contrario es el de la obra cerrada. Toda la riqueza del proceso anterior se ve truncada. La obra no permite el diálogo, las interpretaciones, el cuestionamiento: es una imposición a rajatabla. En un caso extremo: la imposición de una forma de conocer, de aproximarse al arte, cuando no de pensar. Ahora bien, serían argumentos muy cuestionables los de que existen obras tan cerradas como para incurrir en todas estas faltas, y obras tan abiertas como para no poder llegar a ninguna conclusión realmente acertada sobre ellas. Existen, sí, obras fascistas, propagandísticas para inspirar una cierta ideología o apoyar campañas y regímenes, así como existen obras cuyo contenido, y cuyo autor también, se resisten a interpretaciones, con finalidades la mayoría de las veces también externas al arte. Las del primer conjunto podrían ya considerarse obras cerradas.

Respecto al segundo conjunto de obras, su calidad nos lleva a pensar en otro método de interpretación por parte de la crítica. Uno que ejerce violencia contra la obra, para hacer igualmente



un uso de esta. Por ejemplo, en una interpretación psicoanalítica un crítico lacaniano podría argumentar la presencia de evidencias de las etapas sexuales infantiles en la obra Picasso, que delaten que el autor no ha podido superar dignamente el complejo de Edipo, lo que trunca su sublimación... Una interpretación de esta índole, que sale de la obra y que invade otras áreas que le son ajena, son, según Eco, no una interpretación, sino un uso de la obra para establecer una idea o un postulado. La interpretación se queda y se ejerce en el texto, el uso de una obra, redundante, usa la obra para decir algo más de otra cosa.

Esta es la teoría de cómo una obra de arte puede ser abierta o cerrada. Ahora bien, la pregunta que Eco se planteaba en *Obra abierta*, pero que sólo pudo encontrar satisfacción verdadera hasta *Lector in fabula*, es la de cómo es posible, con base a qué elementos, una obra se presenta abierta o cerrada. ¿Qué hay en esas figuras, en ese ritmo, en esos personajes del Picasso, qué hay en la ideología de ese personaje de la obra literaria, qué en su uso de determinado tiempo verbal que mantienen abierta, o cierran la obra? En suma, no es otra la pregunta, planteada exhaustivamente, que la de qué es una obra cerrada y qué una obra abierta. Para contestarla, habría que enfocarnos en *Lector in fabula*.

En términos semióticos, lo que Eco postula en *Lector in fabula* es el proceso por el cual el lector transforma *expresiones* (las formas de los signos, es decir, lo que ya dijimos antes: ritmo, actantes, contextos, tiempos verbales, etc.) en *contenidos* (significado). Esta visión es estrictamente semiótica, una puntualización de la conceptualización de Peirce sobre la cualidad triádica del signo y la semiosis.

Ahora bien, si hemos mencionado antes que el lector es solo una de las piezas fundamentales del juego de la literatura, es porque entendemos las varias críticas que se pueden extender a las decisiones teóricas y metodológicas de esta tesis. Ciertamente una crítica literaria



del siglo XXI poco podría apelar al regreso a estructuralismos y estéticas de la recepción, en tanto que otras críticas, unas materialistas, por ejemplo, consideran centrales al texto y el autor. La discusión que proponemos aquí poco tiene que ver realmente con la validación de una u otra corriente de la teoría literaria, pero nos permitimos en este punto la facultad de mostrarnos críticos con la metodología en tanto que nos permite justificarla.

Lector in fabula deja claro en sus páginas iniciales que su base teórica principal es la semiótica de Charles Sanders Peirce, en especial por lo que a la lectura triádica de un signo respecta. Un semema “es un texto virtual” (41), se menciona, en tanto que un semema, o signo, despliega, con su lectura, una galaxia, una enciclopedia de términos y significados ligados a él, que lo interpretan y que refieren al objeto al que hace referencia. Esta cadena potencialmente infinita de interpretaciones para el signo es lo que se denomina como semiosis infinita, y solo puede ser truncada cuando el lector llega a una interpretación satisfactoria (lo que señalamos párrafos antes: cuando se alcanza una conclusión interesante, el “influjo” comunicativo de la obra se sosiega, pero no para siempre) desde el punto de vista del pragmatismo peirceano, es decir, cuando esta interpretación modifica, de alguna manera, su realidad, en un sentido cognoscitivo o epistemológico. Este manifiesto interés por la interpretación del lector automáticamente posiciona a Eco y su teoría dentro del marco de la estética de la recepción, una abstracción tan válida como reduccionista.

La estética y la teoría de la recepción responden a esa falta de atención al lector de la que medianamente “pecaron” teóricos formalistas rusos y otras corrientes de inicios del siglo XX, como el *New Criticism*. Así, los estudios sobre cómo el lector contribuye a construir la obra, toman fuerza de la mano de Wolfgang Iser y Hans Robert Jauss; pero también es común ver en manuales de teoría literaria, por ejemplo, a Eco en esta sección antes incluso que en los apartados de



estructuralismo. El ejemplo por antonomasia es el libro de Eagleton, *Una introducción a la teoría literaria*. De aquí tomaremos únicamente una noción que nos interesa: la crítica dirigida a poner un peso demasiado estricto y unívoco en el lector.

Más adelante hemos de asegurar con argumentos más incisivos que existen repercusiones para lo que ChatGPT y usuario crean y extienden y denominan literario. La consideración para esto es que la literatura, como hemos dicho en el planteamiento del problema, tiene la capacidad de influirnos, de afectarnos. Nosotros podemos decidir el mostrarnos pasivos ante la literatura. La crítica que erige Eagleton se dirige a este pensamiento, particularmente desde la visión liberal humanista de Iser, y aunque esta no es precisamente compartida por Umberto Eco en el *Lector in fabula*, la recuperamos aquí como un acto de atención en favor de formular un criterio más adelante, uno que nos prevenga de acometer observaciones demasiado ingenuas y reduccionistas sobre la función del lector en la literatura y su apertura a los textos.

Eagleton extiende una crítica a una noción endeble en el liberalismo humanista de Iser al mencionar que:

para que el texto nos transforme es preciso, ante todo, que nuestras convicciones tengan un carácter bastante provisional. Sólo puede ser un buen lector quien *ya* es —de antemano— liberal: el acto de leer produce un tipo de sujeto humano que ya se da por descontado (101).

Y esta no es otra que una visión de un lector empírico demasiado ideal, una abstracción de un lector que de hecho parece bastante competente para la literatura, por no decir que Eagleton refiere a que Iser intenta postular lectores demasiado influenciables. Pero esto es, precisamente, eso, una idealización del lector, de la situación del receptor del texto.

La clase de lector a quien la literatura va a afectar más profundamente está de antemano equipado con capacidades y respuestas del tipo “adecuado”; es hábil en el manejo de ciertas



técnicas de crítica y muy capaz de reconocer ciertos recursos literarios convencionales.

Ahora bien, éste es, precisamente, la clase de lector que mucho menos necesita ser afectada (101).

Esta visión crítica es válida ante todo para la teoría de Iser, pero no escapa a una visión en conjunto que puede ser propia de una estética de la recepción literaria. Una visión que apele indirectamente a un trato funcionalista de la literatura, o a extenuar la relación texto-lector a un duelo intemperante de interpretaciones. Eagleton continúa ampliando su lista de críticas a los teóricos de la recepción (102, 103, 105), críticas que incluyen incluso a Roland Barthes y su visión del placer del texto.

El pensamiento de Iser no es el de Eco. Más correcto sería enmarcar a este dentro de la semiótica misma de Peirce, aunque el autor esté dispuesto también a hacer una revisión incluso de las bases de Peirce. Lo que sí está claro es la defensa que Eco hace por la *libertad interpretativa*, axioma que acompañó su obra hasta el final. Esto es lo que lo posiciona, de una manera u otra, dentro de la estética de la recepción. Por eso ha sido necesario este recorrido previo al abordaje de los conceptos, como una especie de prevención, como una toma de conciencia de que el lector es, reiteramos, solo uno de los papeles fundamentales de la literatura. Ni los mismos autores que se aferren a que el efecto de su obra pueda commover a ciertos grados ideales a su lector estarían exentos de someterse a una revisión crítica de su obra desde otras perspectivas. Tal es el caso de Borges, por ejemplo.

Pero en este punto nos hallamos cercanos ya a la apertura de la caja de Pandora que es la teoría literaria en sus entrecrucos y límites difuminados entre corrientes y escuelas literarias. A la toma de conciencia de las críticas mencionadas, sugerimos también que un factor positivo de la metodología de Eco en el *Lector in fabula* está dado por su incursión en los estudios de



comunicación, factor que traslada a esta obra al tener en cuenta los códigos individuales de los lectores empíricos que permiten el intercambio informativo con una obra. Por otro lado, una prevención que hemos tomado para esta investigación está dada por la metodología, donde si bien partimos de una suposición —una necesaria para el experimento— sobre el lector empírico del cuento que se creará en conjunto con ChatGPT4, esta encuentra también sus fundamentos en los modelos que hemos seleccionado para la construcción del cuento.

Dejando la teoría literaria y las críticas de Eagleton de lado, revisaremos ahora de lleno la teoría de Umberto Eco en *Lector in fabula*. La lógica de la metodología de Eco demanda en algún punto llegar a conclusiones sobre las partes involucradas en la obra de arte: el texto, el autor del texto, y el lector del texto. La asignación de papeles a cada uno permite observar y definir el proceso de interpretación que el lector hace de la obra, es decir, permite entender cómo esas expresiones se transforman en contenidos, cómo se crea el significado. Esta visión lleva al semiólogo a reflexionar sobre el papel del lector como el dador de sentido del texto en tanto que completa su incompletitud. Y para un marco teórico que aborde el intercambio semiótico entre un texto de una IAG y un lector humano, habremos de comprender más a profundidad el papel que el lector tiene con el texto, hecho que nos revela, por supuesto, más sobre el mismo texto también, lo cual es lo mismo que decir que nos abre un panorama más amplio sobre la comprensión e interpretación de una obra de arte, un elemento clave en esta investigación.

El papel del lector

Si se ha dicho antes que en la teoría de Peirce un signo es un texto en potencia, un texto virtual, esto es válido también a un nivel macro. Eco traslada esta noción al texto literario con la hipótesis de que un texto está incompleto (73) y el lector debe completarlo. Un texto es “un



mecanismo perezoso (o económico) que vive de la plusvalía de sentido que el destinatario introduce en él” (76). Es lo ya dicho en términos estrictamente semióticos: una expresión necesita ser puesta en relación con su contenido, y es el lector del signo, el lector del texto, quien pone en marcha este procedimiento de correlación. Esta es la visión que lleva más adelante a Eco a formular específicamente el concepto de Lector Modelo, y a establecer también otros conceptos, como el del Autor y los de obra abierta/cerrada.

Antes de abordarlos, surgen otro par de consideraciones importantes, que habremos de revisar también en ChatGPT4 y su cuento. Si un texto espera de un lector ser completado, es porque *el texto prevé a su lector*. Ahora bien, la capacidad de un texto para prever a un lector desemboca precisamente en el Lector Modelo. Más adelante aclararemos que el Lector Modelo, en pocas palabras, es la idea de cómo le gustaría al texto ser leído. Esta estrategia implica, por supuesto, un trabajo semiótico. Sería una falla demasiado ingenua por parte del autor de un texto considerar, como ya vimos, una idea de lector preconcebida, un estereotipo. De modo que Eco lo que propone es buscar cómo un texto contribuye, más que a preverlo, a *crear* a su lector, cómo “postula una competencia gramatical por parte del destinatario” (74). Esta cuestión se aborda cuando se entiende que el texto pone de cara una *competencia enciclopédica* la competencia enciclopédica que ya posee el lector, lo cual es su forma de proporcionar una suerte de “manual de lectura”.

La enciclopedia es un término similar al de código, ambos presentes también en toda la obra del semiólogo por ser inherentes a la teoría semiótica. La enciclopedia es el conjunto de todas las interpretaciones de un signo, un conjunto no describible en su totalidad (Eco, *Semiótica y filosofía del lenguaje* 133). La visión enciclopédica de un término forma parte de lo que se conoce como semántica composicional⁵⁷, dado que al ser la enciclopedia un conjunto de interpretaciones,

⁵⁷ En otros textos de la obra semiótica de Eco, el término “semántica instruccional” se revaloriza por encima del de “semántica composicional”, aun si ambos se adaptan a lo que para el autor es el concepto central de enciclopedia, es



estas interpretaciones derivan no sólo en más términos de valor semántico, sino también en aquellos que tienen un peso pragmático, estableciendo así recorridos de significado fructíferos y amplios (150), basados no sólo en la semántica de corte de diccionario, es decir, de términos aislados. Pero el modelo de diccionario como modelo de comprensión del significado no es, en general, “una condición estable de los universos semánticos” (158), aunque eso no invalida este modelo, del que pueden valerse disciplinas específicas. En un lenguaje más ameno, el autor explica el término de enciclopedia en *Seis paseos por los bosques narrativos*:

He aprendido que para muchas cosas, en el pasado, he podido fiarme del saber ajeno, reservo mis dudas para algún sector especializado del saber, y para lo demás me fío de la Enciclopedia. Por Enciclopedia me refiero a un saber maximal, del que poseo sólo una parte, pero al que eventualmente podría acceder porque este saber constituye una especie de inmensa biblioteca compuesta por todas las enciclopedias y libros del mundo y por todas las colecciones de periódicos o documentos manuscritos de todos los siglos, incluidos los jeroglíficos de las pirámides y las inscripciones en caracteres cuneiformes. La experiencia y una serie de actos de confianza con respecto a la comunidad humana me han convencido de que lo que la Enciclopedia Maximal describe (no raramente con algunas contradicciones) representa una imagen satisfactoria de lo que llamo el mundo real. Pero lo que quiero decir es que la manera en la que aceptamos la representación del mundo real no difiere de la manera en que aceptamos la representación del mundo posible representado en un libro de ficción (99).

dejar, un modelo semántico amplio que considere una serie de situaciones y contextos también pragmáticos, a diferencia de como sucede con un modelo semántico de diccionario, como se verá a continuación. Más adelante, durante la discusión de los mecanismos semánticos y de conocimiento de los Grandes Modelos de Lenguaje (capítulo 2), la distinción entre semántica instruccional y componencial tendrá un peso mayor.



Si el papel del lector está dado como el que pone en marcha la máquina perezosa, es precisamente porque una enciclopedia nunca logra estar completa, como ya también dijimos. Esta misma noción es aplicable para cuando el texto intenta altruistamente proporcionar su enciclopedia a su Lector Modelo: una forma de invitar a leerle, una liberación de las claves para hacerlo, pero no todas las claves.

...todo intérprete que deba interpretar un texto no está obligado a conocer la enciclopedia completa sino sólo el fragmento de enciclopedia necesario para comprensión de dicho texto. Una semiótica textual también estudia las reglas que el intérprete, sobre la base de ‘señales’ contendidas en el texto (e incluso sobre la base de un conocimiento previo), usa para decidir cuál es el formato de la competencia enciclopédica que se necesita para abordar ese texto. Así se establece también una distinción entre *interpretación* de un texto y un uso indiscriminado del mismo (*Semiótica y filosofía del lenguaje* 134).

Este último fragmento hace referencia al ejemplo que expusimos sobre el Guernica, es decir, que un lector puede hacer una interpretación sobre el texto sobre la base misma del texto, sobre lo que puede leer en él. Una salida de este para postular un hecho o una idea sobre el texto se torna “violenta”, y deriva en un uso de este en lugar de una interpretación.

Luz Aurora Pimentel reconoce en *El relato en perspectiva*, la importancia de la competitividad para la lectura, de modo que agregamos con su teoría que, implícita en las deducciones que giran en torno al lector, el proceso de lectura no es uno pasivo: “Leer es participar en la construcción del texto y del mundo porque la literatura, y en especial la narrativa, es parte del mundo de la acción sólo cuando es leída. Es por ello que importa sobre manera *aprender a leer...*” (163). Pero lo justo es equilibrar esta sentencia volviendo a que la dinámica entre lector y texto no debe tornarse la de una tarea o una imposición. A no ser que estemos frente a una literatura



(neo)vanguardista, el texto, como hemos expuesto con el concepto de enciclopedia, determinará ciertas pautas para su lector. El papel del lector es aceptarlas, voluntariamente y haciendo uso de todas sus capacidades y disposición, y entonces seguir el rastro del texto, rastro que se condensa de mejor manera en el concepto, precisamente, de Autor Modelo. Esta estrategia es una estrategia tanto de construcción textual como de lectura que el lector empírico debe figurarse. Solo así dicho lector alcanzará el grado de Lector Modelo, grado que el texto mismo desea.

Esta consideración de la que parte el *Lector in fabula*, la de que el lector debe poner en correlación las expresiones con sus contenidos, va guiada por la revisión de un modelo narrativo, modelo que incluye un conglomerado de estructuras presentes en el texto: el tema, los *frames*, macroproposiciones, códigos del lector, circunstancias de enunciación, etc. En lo que respecta al cuerpo de la tesis, estaremos echando mano de algunos de estos conceptos, pero lo que más nos importa es poder definir con claridad las estrategias del LM y el AM implícitas en el texto generado con ChatGPT4. De modo que lo siguiente es abordar directamente dichas categorías.

El autor empírico

El autor empírico es la figura humana-histórica, nominal, que construye el texto. Es posible que aquí podamos ceñirnos a Barthes con la muerte del autor y restar interés a la figura del autor empírico. Sin embargo, para la propuesta que se delineea en los objetivos, es imperativo también contar con información sobre el lector empírico, en especial cuando esta es, potencialmente, una inteligencia artificial. Las conclusiones que de esto podamos extraer son importantes más allá del contexto de esta tesis, como estableceremos en el apartado 3.3.

El semiólogo define al autor empírico de manera que este configura, como es obvio, los contenidos de los textos. La figura empírica, histórica, puede “manifestarse” a través de un estilo,



como papel actancial (sujeto de un enunciado), como sujeto ajeno al enunciado pero presente “en el tejido textual más amplio” (88). Esto último quiere decir más propiamente, “como la generación de una estrategia”, que se formula luego como el LM:

el autor empírico, en cuanto sujeto de la enunciación textual, formula una hipótesis de Lector Modelo y, al traducirla al lenguaje de su propia estrategia, se caracteriza a sí mismo en cuanto sujeto del enunciado, con un lenguaje igualmente "estratégico", como modo de operación textual (89-90).

Lo mismo que con la estética de la recepción, diferentes pensadores dan diferente valor y significado al autor empírico. Nosotros lo consideramos importante en tanto que el lector empírico no es ajeno a la idea del autor empírico, y este podría rescatar importantes nociones del mismo a la hora de enfrentarse al texto, nociones como la época en que vivió el autor o nociones generales de su poética, elemento extratextual. Si el autor empírico de los textos fuese una IA, ¿qué consideraciones se activarían? Es lo que se discute en el primer capítulo de esta tesis.

Por otro lado, el mismo texto, en su condición de proveer las competencias necesarias para acceder a él, podría abrir la ventana al autor empírico directamente si el texto se muestra con la etiqueta de novela autobiográfica, por ejemplo. Siempre es posible que el lector confunda la voz del texto, o sea, el AM, con el autor empírico: “con frecuencia, el lector empírico tiende a rebajarlo [al AM] al plano de las informaciones que ya posee acerca del autor empírico como sujeto de la enunciación” (90). Esta consideración no es ociosa porque es válido hacer inferencias extratextuales, unas más válidas que otras, por supuesto⁵⁸. Un cuento de un escritor bien conocido y cuyo estilo está fuertemente codificado en la cultura, hará que el lector empírico *no tenga que*

⁵⁸ Cuando Pimentel habla del “autor implícito” menciona lo mismo que Eco aquí: “...con frecuencia es una imagen que no escapa a la contaminación por el conocimiento de datos biográficos [del autor] y/o anécdotas más o menos apócrifas” (174).



partir desde cero para el proceso de interpretación: su proceso de interpretación tendrá ya ciertas condiciones, así como lo hemos señalado antes en el apartado del autor empírico. “La configuración del Autor Modelo depende de determinadas huellas textuales, pero también involucra al universo que está detrás del texto, detrás del destinatario y, probablemente, también ante el texto y ante el proceso de cooperación” (95). Este acontecimiento feliz nos permitirá volver sobre las huellas también del autor empírico, especialmente sobre el funcionamiento clave de ChatGPT, de su Transformador, para delinear conclusiones fructíferas sobre el porqué de este proceso de cooperación interpretativa.

El lector empírico

La literatura respecto al lector empírico es, por obvias razones, más escueta que la del resto de categorías. La idea de un lector empírico opera casi como una presencia sine qua non, pero es complicado entrar en una definición operativa lo suficientemente sólida y que no se comprometa con alguna otra disciplina. Es decir, un lector empírico puede encontrar una definición desde, digamos, el psicoanálisis. Lo más razonable para el caso de esta investigación, es ceñirnos a lo que Eco propone en su modelo comunicativo y narrativo, y considerarlo solo como las posibilidades que conforman a los lectores empírico, a saber: que el lector empírico cuenta con una enciclopedia que difiere de la del autor empírico, y por ello es necesario que el texto postule su propia comprensión, si bien no exhaustiva, por supuesto; que el lector empírico cuenta con sus propios códigos ideológicos que interferirán al momento de la actualización del contenido de la obra; que hay una serie de circunstancias de enunciación que le interesan al lector empírico, tales como informaciones generales sobre el emisor del texto, época, contexto social, etc. (*Lector in fabula* 106-122).



A manera de referencia bibliográfica: un autor que ha dedicado un estudio literario a entender mejor al lector empírico es Norman Hollando, con su libro de 1975 *The Dynamics of Literary Response*.

El Autor Modelo

Como ya se dijo, Autor Modelo y Lector Modelo son estrategias textuales (89) en la creación de un texto, que permitirían al lector empírico llegar a una, o varias, interpretaciones. El autor empírico se caracteriza a sí mismo como Autor Modelo a manera de estrategia cuando busca crear una hipótesis de Lector. Esto lo hace a través de un lenguaje “igualmente estratégico como modo de operación textual” (90). Cabe aclarar que un autor empírico no piensa en términos de estos conceptos literarios, es decir, no piensa en formularse un LM o no planea cómo formarse él mismo como AM: simplemente tiene una idea de *cómo quisiera que su texto fuera leído*, y la forma en que esta idea se visibiliza es a través de las diferentes decisiones que se toman a lo largo de un texto, decisiones estilísticas y poéticas, que desembocan en el discurso narrativo. Sin embargo, quien termina de formular la hipótesis que representa el AM, *es el lector empírico*. Es el lector empírico quien tiene que, de cierta manera, “descifrar” este Autor Modelo. Hermenéuticamente se diría que debe dejar hablar al texto, escucharlo. Formular un Autor Modelo es ya una “tirada interpretativa”. Es el lector empírico quien se enfrente a las palabras, a las estructuras (los personajes, los acontecimientos y los temas) de una obra, y las interpreta, da contenido a las expresiones, para figurarse esa estrategia textual llamada Autor Modelo. Es decir, el lector empírico es quien “encuentra” esta estrategia guía en el texto:

... el lector empírico, como sujeto concreto de los actos de cooperación, debe fabricarse una hipótesis de Autor, deduciéndola precisamente de los datos de la estrategia textual. La



hipótesis que formula el lector empírico acerca de su Autor modelo parece más segura que la que formula el autor empírico acerca de su Lector Modelo (90).

De modo que aquí empiezan a delinearse los medios para acceder a una interpretación a través del proceso de lectura. El lector más óptimo, que acepta este “trato”, esta guía, de su Autor Modelo, aun con más o menos reservas, es el Lector Modelo.

El Lector Modelo

La conceptualización de la cooperación interpretativa según los planteamientos de Eco solo puede estar completa una vez que se han definido Autor Modelo y Lector Modelo, y también sus “versiones” empíricas. Por lo tanto, resta definir lo que es el LM como estrategia textual:

Para organizar su estrategia textual, un autor debe referirse a una serie de competencias (expresión más amplia que ‘conocimiento de los códigos’) capaces de dar contenido a las expresiones que utiliza. Debe suponer que el conjunto de competencias a que se refiere es el mismo al que se refiere su lector. Por consiguiente, deberá prever un Lector Modelo capaz de cooperar en la actualización textual de la manera prevista por él y de moverse interpretativamente, igual que él se ha movido generativamente (80).

Los medios a través de los cuales esta previsión se completa son un conglomerado de técnicas tanto lingüísticas como estructurales: “la elección de una lengua ..., la elección de un tipo de enciclopedia ..., la elección de determinado patrimonio léxico y estilístico... Puedo proporcionar ciertas marcas distintivas de género que seleccionan la audiencia ... puedo restringir el campo geográfico...”, etcétera⁵⁹ (80). De este modo, lo que Eco propone es que el autor

⁵⁹ Ahondar más profundamente en los medios del autor nos llevaría a la formulación del estilo de un texto, y una vez obtenido esto, pero ampliando las muestras, se podría hablar del estilo del autor, y, si se agregaran a estas confirmaciones estructurales consideraciones extratextuales no sólo sobre su obra propia, sino sobre su aprehensión y visión de la literatura, además de otras marcas genéricas de la época, se dará con la poética del autor.



presupone ciertas marcas que configurarían la competencia de su lector, como ya explicamos con el término de enciclopedia, que el *autor instituye* (81): “De manera que prever el correspondiente Lector Modelo no significa sólo ‘esperar’ que éste exista, sino también mover el texto para construirlo. Un texto no sólo se apoya sobre una competencia: también contribuye a producirla” (81).

Por eso la visión del LM no puede desconectarse del AM: porque las previsiones interpretativas del lector empírico para formular el AM del texto definitivamente están conectadas a estos medios, a estas marcas que el autor empírico ha dejado para su LM. Así, guiado por el influjo del Autor Modelo, el lector que se encuentre dispuesto a la cooperación deberá seguir una serie de “huellas” en el texto, y seguir el camino, o los caminos, las disyunciones, las vías que el AM del texto le está proponiendo. Vías que —a riesgo de ser repetitivos— fueron programadas para el LM. El Lector Modelo es, pues, el *lector dispuesto a seguir estas huellas del texto*, aun si estas se abren como un dédalo, aun si estas no le llevan a la salida que esperaba, o le ofrece una multiplicidad de salidas del texto. El Lector Modelo es el tipo de lector que se abre a absorber la competencia que postula el texto, que decide “seguir el juego” a este. Y así, en esta hermenéutica, es como puede el lector postular una interpretación sobre el texto, acción que más propiamente delinearemos a continuación, y que completará a los conceptos hasta ahora expuestos.

Concepto	Descripción	Acciones
Autor Modelo (AM)	Estrategia textual interpretada por el lector empírico del texto. Se formula en el texto a través de un lenguaje estratégico al que recurre el autor empírico, en el cual involucra su estilo y poética. Este posicionamiento estratégico está enlazado a la formulación de un Lector Modelo	Entablar diálogo con su Lector Modelo, proporcionándole la enciclopedia adecuada para la lectura de la obra y “guiándole” a través de esta a la posibilidad de entablar interpretaciones

Autor empírico	Figura humana-histórica, nominal, que construye el texto. Se manifiesta de diversas maneras en el este: como un estilo, como sujeto de la enunciación mediante un lenguaje “estratégico”. Consideraciones contextuales y culturales sobre este autor también pueden ser consideradas por el lector empírico al momento de formular interpretaciones.	Formular una hipótesis del Lector Modelo de su texto Construir el texto (estructura) y el discurso (cadena sintagmática de sentido)
Lector Modelo (LM)	Estrategia textual de lectura prevista por el autor empírico del texto. La visión de un LM se prevé desde la lengua en que está escrita una obra, así como por el tipo de enciclopedia que ha de postular el autor empírico que permita el conocimiento y recorrido por su obra. Es lo que en otras teorías se denomina “lector ideal”. Es una visión ideal de cómo al texto “le gustaría” ser leído.	Entrar en diálogo con el Autor Modelo al interpretar el texto con base a las propuestas y marcas encontradas en el texto (enciclopedia).
Lector empírico	Lector del texto. Contribuye a llenar los “huecos” del texto con su propio bagaje cultural e ideológico, y mediante su propia competencia enciclopédica, la cual también dispone a ser modificada por la del texto en su condición de metáfora epistemológica.	Formular hipótesis de Autor Modelo (estrategia textual). Interpretar el texto: correlacionar expresiones y contenidos.

Cuadro 2

La cooperación interpretativa en el texto narrativo: obra abierta, obra cerrada

Las estrategias textuales y los participantes del texto han sido conceptualizados. Para retomar ahora la dinámica de la interpretación, tenemos que remitirnos a su aplicación dentro del texto, lo cual nos lleva a fijar la mirada también en la estructura del mismo, estructura que, como ha quedado dicho, está enmarcada por la estrategia que es el Autor Modelo, generada a través de la previsión de un Lector Modelo. Esta labor desembocará, pues, en un factor importante: en la calidad de la cooperación interpretativa que expone el texto. En este factor se encuentra lo que Eco, como hemos dicho, se figuraba lograr ya desde *Obra abierta*: buscar qué es lo que permite la



apertura interpretativa de la obra. En el *Lector in fabula*, se estudia a fondo la consideración de que la obra puede también, en su defecto, ser cerrada en este intercambio entre texto y lector. De cualquier modo, para encontrar esta calidad en la cooperación interpretativa, según la metodología de la semiótica textual, deben definirse primero AM y LM.

El papel del lector es completar un texto. Eco utiliza la metáfora de la “máquina perezosa” (39) en el *Lector in fabula* para referir a cómo el texto debe ser puesto en marcha por su lector. Lo que hay que mencionar respecto a esto ahora, es que un acto de lectura es de entrada un acto de interpretación, pero uno que va tomando forma cuando se avanza linealmente por el texto. Esta misma noción es la que acompaña a la lectura de un signo cualquiera. En *Semiótica y filosofía del lenguaje*, Eco busca rescatar el concepto de signo de una visión que lo encarcela en relaciones de semejanza o de reducción y lo propone como algo que amplía su propia comprensión, por la facultad de que “... la característica principal del signo consiste en su capacidad de estimular interpretaciones” (71). Así, pues, no se trata únicamente de un juego de sustituciones donde un signo remite a otro signo y ahí termina. Al contrario: “... todo interpretante⁶⁰ ... no sólo retraduce el ‘objeto inmediato’ o contenido del signo, sino que amplía su comprensión” (71). Este mismo proceso lo enuncia Eco cuando en el *Lector in fabula* asegura que “el semema debe aparecer como un texto virtual, y el texto no es más que la expansión de un semema” (37), lo cual no nos dice sino que el texto, como el signo, puede estar abierto a interpretaciones. Y esto es válido para todo texto, aun con la existencia de los criterios de obra cerrada o abierta, que en un punto comienzan a operar de manera más bien estética. De momento sólo cabe terminar de postular por qué todo texto invita a su lector a hacer interpretaciones.

⁶⁰ Un signo que interpreta un signo que es, a su vez, un signo en sí mismo.



Si ya establecimos, pues, que dar lectura a un texto es emitir un acto interpretativo, lo que hay que tener en cuenta para este ejercicio es que la lectura de una obra es lo que la dota de vida. Y esta lectura lleva un camino definido por el Autor Modelo, que guía a su Lector Modelo. Este intercambio es el proceso de interpretación, que puede ser basto y, generalmente, abierto. En el contexto del libro de *Lector in Fabula*, es decir, en el contexto que usamos aquí al delinejar las estrategias textuales, se entiende interpretación como “... la actualización semántica de lo que el texto, como estrategia, quiere decir con la cooperación de su Lector Modelo” (252). Y si el texto aquí es mencionado como estrategia, habría que enunciar lo que es la interpretación también en término de sus dos estrategias textuales principales. Dice Eco: “...la noción de interpretación supone siempre una dialéctica entre la estrategia del autor [AM] y la respuesta del Lector Modelo” (86). En virtud de esto, leer un libro, un texto, un poema, un cuento, es, pues, emitir interpretaciones sobre el mismo, que van desde postulados básicos que estructuran el texto: ¿de qué trata? ¿quién es el protagonista?, siguiendo con ello las indicaciones del AM, hasta interpretaciones más profundas y pragmáticas (“¿cuál es el mensaje de la obra?”, por ejemplo). La cooperación interpretativa que emite el texto, o que, por otro lado, deniega, es una parte fundamental de la lectura. Semióticamente, ya se ha dicho, leer un signo es interpretarlo. Leer un texto conlleva también interpretarlo, darle forma, hacerlo hablar, a la manera de Gadamer. No es un proceso que deba dejar indiferente al lector, y es por eso que autores como Eco consideran la apertura de una obra, es decir, la abierta invitación de la obra a que su lector la explore, como esta metáfora del mundo que es, una cuestión insoslayable. Es lo que hemos referido en el planteamiento del problema: la libertad interpretativa depende de que las obras la estimulen también.



Podemos refinar un poco esto de la libertad a establecer interpretaciones y lecturas ahora en palabras de la teórica Luz Aurora Pimentel. Ya aludimos antes a que las lecturas no pueden ser infinitas. El texto también cuenta con autonomía, y no deposita todo su peso en la carga subjetiva del lector. Pensar en los polos extremos del proceso de la cooperación interpretativa nos puede revelar un poco más de la cuestión que intentamos dilucidar. Reafirma Pimentel al respecto que el lector empírico, lejos de tener esa libertad de abrir la lectura *ad infinitum*, puede ser tomado en términos de una perspectiva relacional, donde este debe poner en marcha el texto, lejos de dominarlo.

La lectura entonces es una *relación* de colaboración entre un “texto perezoso” y un lector “activo” ... Si esto fuera así, si cada texto tuviera un potencial infinito de lectura, todos los textos serían el mismo texto. De la misma manera, la perspectiva relacional podrá contrarrestar, en el otro extremo, la abusiva dominación del texto como “verdad” única y absoluta que el lector sólo tiene que “descubrir” (163-164).

Esta última parte del postulado que considera la situación donde la obra establecer su verdad única y absoluta es, en la terminología de Eco, una obra cerrada.

Podemos ser todavía más incisivos en el ejercicio de la cooperación textual. Otra forma de establecer la cooperación interpretativa es con otro rasgo que se ha señalado en el apartado del Autor Modelo y que remite a Pimentel. La autora menciona que el narrador del relato contribuye a la calidad de la información administrada al lector, misma con la que tendrá que hacer su juego de inferencias. Es decir, el narrador, su forma, es una de las partes de la estructura que contribuye a la apertura o a la cerradura del texto. Del tipo de narrador de focalización libre dice Luz Aurora Pimentel:



le da tanta información al lector que éste tiene la ilusión de recibirla completa (aunque, como ya lo hemos advertido, ni la más omnisciente de las narraciones cuenta todo; siempre hay, aun cuando sea mínima, una labor de construcción y de reconstrucción por parte del lector) (172).

De manera que otra de las formas de la apertura del texto que analizaremos próximamente debe depender de una atención minuciosa al discurso del narrador y a la selección del mismo.

Ahora retomaremos los conceptos principales de la cooperación interpretativa. “Nada más abierto que un texto cerrado” (83), dice Eco sobre el carácter de estas obras. “...aparentan requerir nuestra cooperación, pero subrepticiamente siguen atendiendo sus propios asuntos; son textos ‘cerrados’ y represivos” (304). Pero esto no implica sino que un lector común puede hacer el uso que quiera de estos cuentos. Pero esto es ya, justamente, jugar con el texto, abrirlo. Porque en un texto cerrado,

... una, y sólo una, será la hipótesis correcta. A medida que la *fabula* [historia] se va realizando y se va disponiendo a lo largo de su eje temporal, pone a prueba las anticipaciones, excluye las que no corresponden al estado de cosas al que desea referirse y, por último, traza una especie de línea cosmológica continua según la cual lo que ha acontecido ha acontecido y lo que no ha acontecido ya carece de importancia ... este tipo de *fabula* es cerrada, por cuanto no permite (al final) alternativa alguna y elimina el vértigo de las posibilidades. El mundo (de la *fabula*) es el que es (170).

En términos de la poética del arte contemporáneo, lo que nos interesa y atrae es que las obras sean abiertas. No es este prioritariamente un criterio estético. No nos interesa delinejar por qué una obra abierta es más interesante e induce una experiencia estética más enriquecedora (aunque es evidente el por qué). Nos interesa el cuidado de la libertad de la interpretación y el



alejamiento de la represión que pueden ejercer los textos. Y aquí volvemos a atender a las palabras de Eagleton: no se trata tampoco de incurrir en la exageración de que una obra cerrada acusa a su autor de fascismo. La poética de la obra abierta queda fijada por otros muchos factores culturales y científicos que se reflejan también en la vida diaria colectiva, que incluso devienen de la comprensión del mundo como caos e incertidumbre, factores que han planteado rigurosamente ciencias como la mecánica cuántica desde el inicio y a lo largo del siglo pasado hasta el hoy. En el terreno del arte contemporáneo, esto tiene una influencia insoslayable, por supuesto, en ese juego de espejos conocimiento-cultura⁶¹. Así, pues, en palabras de Eco, la base de la obra abierta es precisamente la de aprehender el arte como una *metáfora epistemológica*, un modelo del mundo, cuyo resguardo es el de la libertad interpretativa. He ahí la conexión con los modelos epistemológicos a grande escala que plantean las ciencias duras. En este punto citamos en extenso la argumentación de Eco en *Obra abierta*, similar a lo antes planteado incluso desde el planteamiento del problema, pero que resulta mucho más puntual, y pertinente para este apartado:

En efecto, es siempre arriesgado sostener que la metáfora o el símbolo poético, la realidad sonora o la forma plástica, constituyan instrumentos de conocimiento de lo real más profundos que los instrumentos que presta la lógica. El conocimiento del mundo tiene en la ciencia su canal autorizado, y toda aspiración del artista a ser vidente, aun cuando poéticamente productiva, tiene en sí misma algo de equívoco. El arte, más que *conocer* el mundo, *produce* complementos del mundo, formas autónomas que se añaden a las existentes exhibiendo leyes propias y vida personal. No obstante, toda forma artística puede muy bien verse, si no como sustituto del conocimiento científico, como metáfora

⁶¹ Por otro lado, entiéndase que la ciencia busca generar leyes, pero no así el arte. De ahí que en el término a introducir el concepto de metáfora tiene un peso importante: el arte, la literatura, son formas *indirectas* de conocimiento del mundo, que no tienen intenciones de generar leyes. Esta distinción debe quedar clara en todo momento, para evitar incurrir en el error de que el arte es un modo de conocimiento directo.



epistemológica; es decir, en cada siglo, el modo de estructurar las formas del arte refleja — a guisa de semejanza, de metaforización, de apunte de resolución del concepto en figura — el modo como la ciencia o, sin más, la cultura de la época ven la realidad ... en el universo científico moderno, como en la construcción o en la pintura barroca, todas las partes aparecen dotadas de igual valor y autoridad, y el todo aspira a dilatarse hasta el infinito, no encontrando límite ni freno en ninguna regla ideal del mundo, pero participando de una aspiración general al descubrimiento y al contacto siempre renovado con la realidad (88-89; las itálicas son de Eco).

He ahí la poética del arte contemporáneo: la de la obra abierta, la de la libertad de interpretar, enmarcadas ambas en un universo cuyos límites epistemológicos se han desdibujado, pero que, por lo mismo, ha jugado a favor de la libertad de expresión para el arte.

Para nuestro proceder únicamente debemos tener en cuenta que hay una importante forma de relacionarnos con la literatura, aunque no afirmemos que la literatura es altamente afectiva y uno debe siempre comportarse pasivamente ante sus injerencias y mensajes. Más bien esta forma de relacionarnos con ella debería ser la de la libertad de interpretarla, de manera que la consideramos, como a todo el arte, como un modelo del mundo, con toda su carga epistemológica contenida. A raíz de este postulado, de esta forma de entender una de las particularidades del arte, la pregunta que nos enlaza con la tecnología, con la inteligencia artificial, es la de si estas tecnologías, con todas sus funciones, sesgos e ideologías que la acompañan, permiten esta libertad interpretativa, o hay algo más sobre lo que llevar nuestra atención.

Estos mismos fundamentos de la cooperación interpretativa son cruciales para tener en cuenta al momento de enfrentarnos a este nuevo tipo de obras, compuestas por y/o a través de agentes no humanos. Es lo que nos disponemos a indagar, nuestro aporte.



Conclusiones

Las estrategias textuales que hemos recuperado de Eco y su *Lector in fabula* son conceptos de la semiótica textual, formulados a partir de las bases del pragmatismo peirceano. Poner en marcha la lectura de un texto es poner en marcha las diferentes estrategias semióticas que rigen al texto, es empezar a darle sentido a través de la lectura de sus expresiones, es entablar un diálogo con el mismo. En tanto que un texto entabla un diálogo con su lector, el mundo del lector se abre al mundo del texto: “la lectura implica también una dinámica de resistencia y de apropiación entre un texto que proyecta un universo, esencialmente ajeno, y un lector que pone en riesgo su mundo al intentar apropiarse de otro” (Pimentel, *El relato en perspectiva* 163). Pero esta relación, que concebimos como fenómeno epistemológico, pues la literatura es capaz de formular componentes de modelos del mundo, no debe exagerarse, al grado de incurrir en los “tropiezos” que Eagleton, por ejemplo, señala de Iser, en una idealización del texto, que hace del lector un Sísifo ante la literatura, un esclavo. Esta relación lo que nos determina es una reflexión sobre nuestra relación con el arte en general. Según la teoría de Eco, si la obra es cerrada, la obra impone su visión, y nuestra relación con el arte toma un cariz distinto a si la obra nos permite navegar libremente a través de ella, abriéndonos paso a la formulación de este modelo del mundo, a un aprendizaje, no impuesto ni moralizante ni programado ni fascista, de lo que la obra tiene para decírnos. La verificación de estos fenómenos es lo que habremos de llevar a cabo en conjunto con ChatGPT4, para emitir conclusiones generales sobre las IA en la actividad creativa-literaria y en las repercusiones que de esto deriven hacia uno de los cuadrantes que componen el átomo de la literatura: el lector.

Pero antes, debemos adquirir una muestra con la que trabajar.



Marco metodológico: Modelo para la creación de un cuento corto con ChatGPT4.

Introducción

Para la verificación de nuestros objetivos, crearemos una muestra narrativa utilizando ChatGPT4. De esta manera podemos verificar también una buena parte del proceso de composición, que nos guíe a la satisfacción de otros de los objetivos de la tesis. No obstante, el método seguido para la construcción de este cuento tiene como imperativo no “contaminar” el experimento: es decir, se ha escrito un cuento de un género tal y de un tema tal, sin ahondar en enunciar directa ni indirectamente los conceptos teóricos que nos disponemos a investigar en este artículo, dado que dicha acción podría “acomodar” los resultados a favor de argumentos planteados en esta investigación. Una vez con este cuento, se aplicarán las categorías de la semiótica textual que nos deberían conducir a determinar la calidad de la relación que el texto establece con su lector. Esta calidad no debería determinarse por la mano del investigador, es decir, contaminando el experimento desde su proceso de composición, para otorgar un fallo ora a la apertura de la obra ora su cerrazón. Es por eso que este marco metodológico está determinado a configurar este proceso de creación de un cuento para analizar como sumo cuidado y atención de esta condición.

Dicho de otro modo, se estaría “simulando” la búsqueda de ayuda de una IAG para la creación de un cuento por un usuario no iniciado en la teoría literaria ni en la ingeniería de prompts, pero sí orientado en lo básico del uso adecuado de estos⁶². Esta metodología de crear un cuento y

⁶² Sobre el prompt y la ingeniería de prompts (desde su primera mención, hemos decidido conservar al término en su idioma original por su uso extendido que resulta más identificable): “Un *prompt* es una cadena de texto que el usuario emite a un modelo de lenguaje para que este realice una acción útil. En la creación de prompts, la cadena de prompts del usuario se pasa al modelo de lenguaje, que genera iterativamente tokens condicionados al prompt. De este modo, el prompt crea un contexto que guía a los LLM a generar resultados útiles para alcanzar un objetivo del usuario. El proceso de encontrar prompts eficaces para una tarea se conoce como ingeniería de prompts.” (Jurafsky, Martin, Speech and Language Processing: An Introduction... 243) (En el original: “A prompt is a text string that a user issues



no de tomar una muestra responde también al espacio del que disponemos para la realización de esta investigación. Y si bien la muestra es minúscula ante los materiales disponibles, estaremos trabajando con las mismas bases con que se crean muchos de esos materiales; o sea: a partir de los *métodos modelo* que posiblemente generaron también esta variedad de materiales, ya que, según se conoce de la ingeniería de prompts y del funcionamiento mismo de las IAG, existen bases bien fundamentadas a partir de las cuales formular instrucciones que proporcionar a una IA, y así obtener los resultados deseados de manera óptima.

Esta búsqueda de un apoyo para la creatividad y la productividad son factores que sugerimos que un usuario de internet podría buscar al momento de utilizar una IAG como asistente para la creación literaria, en este caso, la narrativa. Y es a partir de los productos generados que el investigador entonces interviene con todo el andamiaje teórico.

La aplicación de los resultados encontrados en este marco metodológico se puede encontrar en la sección de apéndices, en la conversación mantenida con ChatGPT4 y en el cuento generado por esta.

La semiótica textual

Como ha quedado claro ya en objetivos y marco teórico, la metodología a la que se recurre para el análisis del cuento en cuestión, y así satisfacer los objetivos, es semiótica principalmente. En puntuales ocasiones se recurre a traducir algunos conceptos también a la terminología del posestructuralismo de Luz Aurora Pimentel, cuya base, a su vez, es Gerard Genette. La razón principal es complementar el diálogo entre estas dos disciplinas, no ajenas la una a la otra dentro

to a language model to get the model to do something useful. In prompting, the user's prompt string is passed to the language model, which iteratively generates tokens conditioned on the prompt. Thus the prompt creates a context that guides LLMs to generate useful outputs to achieve some user goal. The process of finding effective prompts for a task is known as prompt engineering”).



del campo más amplio del análisis literario, de modo que se complementen para llegar a los resultados esperados. Por otro lado, algunos de los conceptos de la mano de Luz Aurora Pimentel y sus bases teóricas permiten explicar con más objetividad y economía lo que en la semiótica textual conlleva recorridos más amplios.

Esta no es, por supuesto, la única metodología que es posible emplear para la exploración del objetivo propuesto, ni siquiera para lo que tenga que ver con la figura del lector. Pero al ser la semiótica la disciplina que estudia el signo y, énfasis en esto, *su interpretación*, proporciona categorías específicas y bien construidas para constatar la calidad y forma de estos recorridos de interpretación. De ahí nuestra elección como principal metodología.

Creación de un cuento corto con ChatGPT4: la intervención del usuario (IU)

El cuento en coautoría con ChatGPT4, que se encuentra en el apéndice B, será sobre el que deberemos leer las formas del Lector Modelo y el Autor Modelo y la calidad de la cooperación interpretativa. Pero el trabajo realizado también abre a otras preguntas válidas de ser planteadas. Por ejemplo, ¿cuál es la participación del usuario humano en todo esto, en términos de cantidad y de calidad? En la sección de los autores empíricos abordaremos más a fondo esta primera cuestión emergente. Sin embargo, en este apartado proponemos un método de clasificación: cinco niveles que permitan ubicar de manera satisfactoria la participación del usuario en el proceso creativo al lado de ChatGPT —o cualquier otra IAG—. Este modo de proceder, argumentamos, es crucial, ya que de esta determinación cuantitativa dependen sensiblemente otros factores que tienen relación con la calidad del texto, y con las estrategias textuales a determinar.

Un usuario de internet que se proponga la utilización de las IAG para la creación de textos literarios, académicos, o de discursos, cuenta con varias opciones de intervención (IU: intervención



de usuario). En estos cinco niveles planteados, se establece en una gradación no estricta la cantidad de participación de ambas partes:

Nivel α : donde **IU < IAG**. El peso de la cooperación para la creación recae en la IAG en un grado que varía del 75-99% (aproximadamente) de la totalidad del trabajo elaborado.

Nivel β : donde **IU < IAG**. El peso de la cooperación para la creación recae en la IAG en un grado que varía del 51-74% de la totalidad del trabajo elaborado.

Nivel γ : donde **IU ≈ IAG** (aproximadamente igual).

Nivel δ : donde **IU > IAG**. El peso de la cooperación para la creación recae en la IAG en un grado que varía del 25-49% de la totalidad del trabajo elaborado.

Nivel ε : donde **IU > IAG**. El peso de la cooperación para la creación recae en la IAG en un grado que varía del 1-24% de la totalidad del trabajo elaborado.

De este modo, por ejemplo, el recientemente popularizado caso de la escritora japonesa en que, según sus palabras, utilizó la IA para escribir un 5% de su novela⁶³ nos permitiría ubicar dicha relación de cooperación en un nivel epsilon.

Para el caso que hemos elaborado nosotros en el apéndice B, hemos recurrido a un nivel beta. Las razones para ello se enlazan directamente con que este nivel de participación de la IAG favorece la exposición del proceso de creación del cuento, además de que, por obvias razones, si lo que está en juego es más bien el análisis de las IAG en su labor creativa, un mayor peso debía recaer sobre estas.

Resulta evidente que una verificación de la intervención de ambas partes es difícil de cuantificar de manera estricta, por lo que varias críticas podrían extenderse a este modelo. No

⁶³ Christy Choi, Francesca Annio: "La ganadora de un prestigioso premio literario japonés confirmó que una inteligencia artificial le ayudó a escribir su libro". <https://cnnespanol.cnn.com/2024/01/19/ganadora-premio-literario-japones-inteligencia-artificial-chatgpt-trax>



aseguramos la impermeabilidad de esta clasificación, pero sí *la utilidad de designar la intervención del usuario en el proceso creativo*. Una noción similar es utilizada en el área del *machine learning*, con las nociones de IA supervisada o no supervisada y las diferentes variaciones de cada una. Consideramos que un estudio serio sobre la intervención de las IA en las Humanidades podría valerse de estos conceptos y aun ampliarlos para sus fines específicos. Es lo que hemos llevado a cabo con esta clasificación, que en un futuro será posible refinar.

En el capítulo uno se volverá sobre esta bipartición de la labor creativa, y en el apartado 1.2 se volverá a hacer énfasis en las implicaciones del autor empírico humano, por lo que se volverá a recurrir a esta clasificación.

Nivel	Relación IU - IAG	Factor cuantitativo
α	IU < IAG	IAG interviene en un 75-99% en la elaboración del texto
β	IU < IAG	IAG interviene en un 51-74% en la elaboración del texto
γ	IU ≈ IAG	IAG y usuario intervienen en un 50% aproximadamente
δ	IU > IAG	IAG interviene en un 25-49% en la elaboración del texto
ε	IU > IAG	IAG interviene en un 1-24% en la elaboración del texto

Cuadro 3

Creación de un cuento corto con ChatGPT4: el método

El motivo principal de este apartado metodológico estriba en cómo se ha realizado el cuento que se va a analizar. Crear una muestra sin contaminaciones y sin sesgos que remitan a la ventaja del investigador conllevó una labor minuciosa, la cual se exhibe a continuación.

Para la creación de un cuento a través de ChatGPT4 se ha optado por seguir I) *un modelo de generación de prompts*, y II) *una estructura específica y sistematizada para la generación del*



texto narrativo. La versión utilizada de esta IAG es la más avanzada disponible de forma gratuita en la última mitad del año 2024: Chatgpt4-o.

El modelo y la estructura han sido generados a partir de los siguientes cuatro videos de YouTube: “*Master the Perfect ChatGPT Prompt Formula (in just 8 minutes)!*”, “*The Only 7 Prompts AI Authors Need to Write Their First Book*”, “*How to Write a Short Story in 20 Minutes (ChatGPT + Claude)*”, “*Cómo escribir HISTORIAS, CUENTOS y NOVELAS con ChatGPT ...*”.

La razón global que motiva esta selección tiene que ver con la simulación del actuar de un usuario común de esta IAG. En este mismo ejercicio de la imaginación, podríamos acudir a los ejemplos citados en el planteamiento del problema y en el estado del arte, las obras existentes creadas mediante IAG y que están disponibles en librerías virtuales a la venta. De modo que este hipotético usuario tiene por objetivo algo similar, y no la idea de librarse de algún bloqueo de alguna porción de su obra narrativa, lo cual nos llevaría a recurrir a un nivel inferior de IU (posiblemente delta, como la escritora japonesa), ni buscar mejoras de estilo o de la estructura de su obra (posiblemente épsilon). Como ya mencionamos, este proceder nos haría actuar dentro del nivel que denominamos “beta” (ChatGPT tendrá intervención en un 75 al 99% del resultado final del texto en sí mismo).

Hemos supuesto que un usuario común de internet podría aproximarse a videos de YouTube que sugieran tener las competencias y herramientas necesarias para la generación de estos textos, y por la economía de tiempo que esto sugiere. Ahora bien, las razones específicas para tan puntual selección van desde que los videos cuentan con números elevados de vistas y que provienen de creadores de contenido que han desarrollado previamente otros videos en el tema de las IAG, o en el tema mismo de la escritura creativa a través de estas, como se puede observar en sus cuentas en la plataforma, y que este proceso estructural nombrado con el índice II es el que llevan a cabo



también casos de composición más complejos como los de las mencionadas novelas *Echoes of Atlantis* e *Iris*, ambas en coautoría con ChatGPT4.

No obstante, la pertinencia de este método no nos dejaría satisfechos si no la analizáramos asimismo desde algunos criterios que nos permiten las principales herramientas metodológicas manejadas también en esta tesis, llevadas de la mano con algo del campo de la ingeniería de prompts. Es decir, a través de la lingüística, la semiótica y teoría que conforma el análisis estructural de una obra literaria, verificado a la par con el apartado 12 (“*Model Alignment, Prompting, and In-context Learning*”) del libro *Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition with Language Models*, se ha validado la pertinencia tanto del I) modelo de prompts, y de II) la estructura para la creación del cuento. A continuación, damos cuenta de manera sucinta una primera aproximación.

Hemos verificado, primero, a) que las condiciones que enumera el modelo de la construcción de prompts son fructíferas. Se adecúan a lo que se esperaría de base en un modelo de la comunicación efectiva⁶⁴, donde se incluye un uso favorable del contexto⁶⁵ que enriquece la necesidad de la tarea a llevar a cabo y proporciona a la IA información pertinente de la misma⁶⁶.

⁶⁴ Si bien la comunicación humano-máquina difiere de que involucra dos seres humanos, véase en el *Tratado* de Eco el modelo comunicativo propuesto y que se recupera incluso para el *Lector in fabula* cuya propuesta es la de incluir, a la manera de Jakobson y los estructuralistas coetáneos, el concepto de código como fundamento de la comprensión de la comunicación y que logra ampliar los elementos involucrados en otros modelos más básicos de comunicación.

⁶⁵ Contexto entendido como contexto pragmático, de acuerdo con la propuesta de Teun van Dijk dentro del análisis del discurso (“Conferencia 3: La pragmática del discurso”, *Estructuras y funciones del discurso*): “conjunto de datos a base del cual se puede determinar si los actos de habla son o no *adecuados*” (59; énfasis en el original). Si bien van Dijk sugiere el concepto en un *contexto* de la comunicación social, creemos que la actualización y validación del acto de habla es precisamente lo que opera también para esta comunicación efectiva con la IAG, que le permitirá puntualizar el acto de habla (tarea).

⁶⁶ Un prompt también representa una manera de que la IAG aprenda en términos contextuales. Si bien esta intervención no tiene una influencia en los mecanismos más profundos de la red neuronal operativa de la IAG, sí tiene un efecto en la conversación con el agente humano. De modo que al prompt también se le conoce como “aprendizaje en contexto”, y se pone en marcha a través de la iteración, corrección y refinamiento de los outputs de la IAG por parte del agente humano (Jurafsky, Martin, *Speech and Language Processing: An Introduction*... 242).



Esta tarea, luego, es específica⁶⁷, y en general, con el resto del contenido del enunciado, o prompt, funge como acto de habla⁶⁸. Se indica, así, un proceso comunicativo objetivo y económico. Luego, se ha encontrado como satisfactorio el hecho de que b) los tres videos que implican la construcción de historias con ChatGPT comparten puntos en común, lo que sugiere fuertemente que se ha “descifrado” una base, una estructura básica que aprovecha el funcionamiento de la red neuronal Transformador para esta tarea específica a través de este modelo de indicaciones narrativas⁶⁹. Estos modelos comunes podrían no ser, por supuesto, los únicos, y tal vez ni siquiera los más fructíferos, pero las condiciones enumeradas arriba nos remiten al fallo de que son suficientemente ideales para el ejercicio. Estas son las razones por las que aprobamos, desde estas otras bases teóricas, la selección de estos videos más allá del hecho de la aprobación consensuada que existe entre los usuarios de internet.

Por otro lado, las observaciones indicarían que la acción de solicitarle a una IAG la escritura de un cuento *no tiene que ver tanto con competencias literarias* —que no deja de ser necesario el conocimiento básico de las reglas de la narrativa y un poco de sentido común— *como sí con competencias de la construcción de prompts*. Esto involucra en primera instancia conocimientos básicos sobre el funcionamiento no solo ya de una tecnología GPT, sino de la dinámica de la interacción usuario-asistente, como se deduce en b), proceso que expandiremos más adelante y que exhibe una dinámica que, sin lugar a dudas, por los ejemplos seleccionados, es de continua reescritura, de avance por pasos, discriminación de ideas y pasajes.

⁶⁷ Véanse las máximas de Grice de la pragmática conversacional: máxima de cantidad.

⁶⁸ Entendido este en toda su amplitud también según lo propone van Dijk: un acto de habla es una emisión, y una emisión se usa para realizar acciones (58).

⁶⁹ De acuerdo con Jurafsky y Martin en Speech and Language Processing: An Introduction..., siempre es fructífero comunicarse con un modelo de lenguaje como lo son estas IAG a través de una sistematización de prompts: templates (244, 247), o diagramas, estructuras. Es por eso que este método, llevado de la mano con una buena construcción del prompt, queda validado por observaciones desde la ingeniería de prompts.



Otra razón que nos interesa es que esta aproximación metodológica a la creación del cuento, en efecto, no “contamina” la muestra de la investigación: ninguno de los usuarios de los videos, incluso de los involucrados en *Iris* y *Echoes of Atlantis* hace uso de estrategias de la semiótica textual ni otras disciplinas adyacentes. La muestra estaría contaminada si, a propósito o por ingenuidad, se solicitara a la IA construir un cuento “para luego encontrar su Lector Modelo”, o aun si se le solicitara construir un cuento teniendo en mente cierto tipo de lector, contaminaciones que serían inaceptables. Se evita así, tramposamente, puntualizar los datos que queremos encontrar, que se trata de las estrategias textuales de la IA y, a raíz de ello, el papel que se le da a la cooperación interpretativa con el lector. Se logra, precisamente, evitando tocar estos temas, y manteniendo una aproximación similar a la de este usuario de internet que hemos imaginado, con esta inclinación a aprovechar la rapidez y sorprendente manejo de datos que profesan estas tecnologías para crear cuentos con propósitos comerciales y de divulgación rápida. Tal es la inclinación que parece acompañar a la comunidad de creadores de contenido que se involucran en el tema.

Pasemos ahora al modelo de prompts en sí mismo, conjugado a partir de los videos estudiados.

La propuesta en “*Master the Perfect ChatGPT Prompt Formula (in just 8 minutes)!*” sugiere que un prompt efectivo debería estar construido sobre la base de los siguientes elementos, o bloques: *tarea* (*task*), *contexto*, *ejemplo*, *persona*, *formato* y *tono* (Jeff Su). Se sugiere, luego, una jerarquía de importancia, donde dominan la *tarea*, el *contexto* y el *ejemplo*. La *tarea* es, pues esencial, obligatoria. Un tipo sencillo de tarea se construye con una oración con un verbo al comienzo de esta, conjugado en imperativo (crea, escribe, genera, etc.). Este es el uso sugerido que se tomará, en lugar de apelar por fórmulas como “quiero que crees”, “me gustaría que crearas”,



etc. El *contexto* debería optimizar la *tarea* por medio de ampliar la información a través del trasfondo del sujeto del enunciado, de cómo se vería la *tarea* ejecutada exitosamente y del ambiente que rodea al sujeto. Para el experimento que nos concierne, el contexto lo tomaremos en el sentido que hemos citado de van Dijk, y que en esta propuesta parece exhibirse más únicamente en la contextualización del ambiente y necesidades de un sujeto gramatical o de la enunciación; damos por hecho, al menos para este caso, que el éxito de la *tarea* es imperativo y que estaría de más proporcionar un ejemplo de cómo se vería esta realizada con éxito. En lo que respecta a la construcción de la *tarea*, nos mantendremos apegados a la idea de la máxima de pertinencia en el proceso de la comunicación, que más adelante, durante la ejecución, habremos de concatenar con otros códigos literarios y narrativos. El *ejemplo* es innecesario, pero una guía sí será proporcionada como parte de estos códigos narrativos posteriores, cuando, al momento de ejecutar la orden final que produzca el cuento, se le proporcionará a la IAG la instrucción de dividirlo en capítulos. Este proceder encaja dentro de lo que el autor del video denomina como “estructura de ejemplo”⁷⁰. En cuanto a la *persona*, consistiría en dotar a ChatGPT de cualidades específicas en cuanto a su ejecución, pero discriminaremos este bloque por las posibilidades que existen de que agregue información innecesaria y que contamine el experimento (por ejemplo: “eres una escritora mexicana del siglo XXI”: la cadena de interpretantes podría ser demasiado amplia y diversificarse demasiado. Puntualizarla para evitar esto involucraría dar más información y extender el prompt). La función de la *persona* queda satisfecha a través del *contexto* y el *ejemplo* estructural. El *formato* se traducirá también a términos narrativos: proporcionar el género y una división de capítulos para el cuento es involucrarse ya en lo que el creador de contenido expone por formato. El *tono* será

⁷⁰ Pequeñas proporciones de ejemplos, directos o estructurales, contribuyen al aprendizaje contextual que resulta ser el prompt, de modo que introducir este tipo de ejemplos o señalar ejemplos estructurales a partir de modelos, es también una forma de “enseñar” a la IAG a llevar a cabo la tarea demandada óptimamente (Jurafsky, Martin, Speech and Language Processing: An Introduction... 246-248).



empleado sin pretensión de otros efectos que el de validar el trabajo de la IA, lo cual cumple la función de proporcionarle la información de cómo queremos que opere⁷¹. Se usará cuando en la instrucción final se haga un énfasis optimista en su trabajo.

La II) estructura propuesta, que tiene que ver con el método para escribir cuentos, está dada por los otros tres videos. El método, como mencionábamos, suele ser común entre estos videos, y los tres también reposan ciertas ideas sobre las bases del I) modelo de prompts generales, como el uso del ejemplo, el contexto y aun la persona (por ejemplo: se le pide a ChatGPT4, en el video de “Cómo escribir HISTORIAS, CUENTOS y NOVELAS con ChatGPT (...)”, que escriba como si fuera Philip K. Dick). Así, pues, lo que demuestran estos tres videos en común son dos formas de proceder: la primera, el *brainstorming*. El *brainstorming* consiste en solicitar ideas para una historia y/o para sus personajes, lo cual generará largas listas de argumentos o sinopsis que quedarán guardadas en la memoria del Transformador para el largo de la conversación, y que proporcionarán una mejor contextualización del prompt para la IA⁷². De modo que al *brainstorming* le sigue la *selección, discriminación y expansión* de las ideas iniciales y la formulación de descripciones más concretas. Se logra solicitando, una vez establecido un tema o género directriz, una lista de conflictos, por ejemplo, o de personajes que se adapten a los temas o géneros. Aquí las posibilidades empiezan a bifurcarse entre más intervención tenga el usuario, es decir, entre más elementos decida añadir o corregir durante el ejercicio. Es posible tornar el proceso más interesante si, digamos, se le solicita a la IA una lista de personajes cuyas formaciones morales terminen por encontrarse con uno de los conflictos que habrá de seleccionarse de la lista propuesta.

⁷¹ En 1. Argumentamos que parte de la definición y funcionamiento de las IA tiene que ver con su cualidad de agentes racionales, es decir, agentes dispuestos a llevar a cabo una acción exhibiendo el mejor desempeño y eficacia que la calidad misma de la acción demande, y en medida de las capacidades del agente, claro está.

⁷² Véase lo dicho sobre el prompt como un aprendizaje contextual. A este efecto se le hará referencia también en el capítulo 1 cuando se explore el funcionamiento de las redes neuronales de Transformador y se les defina propiamente.



O se puede comenzar a estructurar la historia, solicitando varios ejemplos de giro de la trama o de clímax y resoluciones, que se muestran a manera de macroproposiciones, según el concepto semántico que también trabaja van Dijk, enseguida seleccionando los más atractivos a criterio del usuario (algo como: “El detective de la historia deberá detener primero a un sospechoso equivocado”, por dar un solo ejemplo). El proceso es en realidad bastante sistemático, si bien puede ampliarse al gusto, y procede por acumulación. La última forma del proceder de este modelo para la construcción narrativa es la de la *corrección*. Es conveniente volver sobre los pasos de la conversación y corregir ideas, ampliarlas, anestesiar cualidades, abrir disyuntivas, agregar, restar, etc.

En síntesis, el proceso se estructura en esta triple formulación: 1) *brainstorming* y *retroalimentación*, 2) *selección*, 3) *iteración*⁷³. Se abre con un *brainstorming*; se seleccionan las ideas más pertinentes y se desarrollan hasta ampliarlas a pequeñas sinopsis o descripciones puntuales, luego, se corrigen o se ajustan las opciones. Aquí es donde se agregan, deliberada o inadvertidamente, otros elementos puntuales, como de estilo, de tono, donde se dota con redondez a los personajes, donde el usuario tiene la oportunidad de escalar de un nivel de intervención beta a uno gamma, por medio de sus propias incursiones poéticas y estilísticas. Se repite entonces el proceso, ya sea con *brainstorming* de subtramas, personajes, espacios, etcétera. El último paso es, por supuesto, solicitar la escritura del cuento con los elementos seleccionados y trabajados y bajo una estructura acordada. De nuevo las variaciones son muchas y abrumadoras: hay usuarios que trabajan en varias salas de chat para evitar mezclas innecesarias de ideas y prompts, por ejemplo.

⁷³ La iteración de correcciones y resultados es similar en tanto que apunta hacia lo que en el área de la ciencia de la IA se conoce como *tunning* o *finetunning*, y que tiene que ver con el refinamiento de aprendizajes basado en la calidad de los outputs en vez de con el texto del pre-entrenamiento alimentado a los Grandes Modelos de Lenguaje (Jurafsky, Martin, Speech and Language Processing: An Introduction... 259).



Para la finalidad de este ejercicio, no es necesario expandirnos demasiado en términos cuantitativos ni cualitativos⁷⁴.

Finalmente, entrarían nuestras consideraciones previas a la formulación de los prompts, donde entran en juego algunos conceptos netamente literarios. Admitimos una definición básica de lo que es un cuento, en oposición a una novela, por ejemplo⁷⁵. Sobre la base de esta idea, entendemos que hay que contar con un género y un tema. El género seleccionado es el policíaco. Esta decisión es de doble naturaleza: tanto por las comodidades como por los desafíos que plantea. Convencionalmente, el género policíaco expone un acontecimiento que requiere explicación, antepone al lector una posible solución que será errónea, y tras de esto una anagnórisis de la solución real. Se está determinando, con esta selección genérica bien codificada en la literatura, una guía para la IA, un *ejemplo* estructural. El desafío que plantea este género es estético, su peso recae en la parte de la anagnórisis y de la corrección de la disyunción de interpretación del lector donde, a manera de ejemplo, “el personaje A no fue el asesino de C, sino el personaje B”.

Considerando todo lo anterior, se optó por una muestra de un par de ejemplos de prompts, con la finalidad de demostrar la validez del método. Este primer prompt es una ejecución efectiva del modelo, con sus elementos más pertinentes, y que economiza la información para optimizar el proceso:

⁷⁴ Otro método de prompting es el “chain-of-thought”, que, como otros métodos de refinamiento de prompts, amplían minuciosamente la instrucción para conseguir *outputs* igualmente más amplias y que sugieran el seguimiento de líneas de pensamiento. Se sabe que se utiliza en el contexto de solución de problemas complejos (Jurafsky, Martin, Speech and Language Processing: An Introduction... 254). El método todavía tendría que ser puesto a prueba para la creatividad, área en que habría que considerar, por supuesto, más factores, empezando por el nivel de intervención humana.

⁷⁵ Desde la literatura entendemos el cuento con base a las definiciones de Cortázar, (“Algunos aspectos del cuento”, página 407); del prólogo al libro *Dos siglos de cuento mexicano*, “Introducción: El cuento: la captación de un momento; de Carlos Pacheco en “Criterios para una conceptualización del cuento”, página 23; y Luis Barrera: “Apuntes para una teoría del cuento”, página 36. Se manejan conceptos de economía, construcción centrada en un tema, un proceder cualitativo y no cuantitativo, que profundice en el tema y no que construya sobre él, la muestra simbólica del tema a través de hechos, personajes, etc.



A) Soy escritor de cuentos de ficción y quiero escribir un cuento del género policíaco.

Proporcioname una lista de cinco ideas que desarrollen el tema de un asesinato misterioso.

Las condiciones para la trama son que primero los personajes que investigan el asesinato deben sospechar que el asesinato ha sido cometido por una entidad fantasmal cuando en realidad fue cometido por una persona.

Si descomponemos el prompt por sus macroproposiciones obtendríamos lo siguiente:

- a) Soy escritor de cuentos de ficción
- b) y quiero escribir un cuento del género policíaco.
- c) Proporcioname una lista de cinco ideas que desarrollen el tema de un asesinato misterioso.
- d) Las condiciones para la trama son que primero los personajes que investigan el asesinato deben sospechar que el asesinato ha sido cometido por una entidad fantasmal cuando en realidad fue cometido por una persona.

Estos elementos se disponen a continuación en la siguiente tabla de prompt A, que explicaría la unión del I) modelo de bloques de prompts y de II) los lineamientos estructurales y de esquematización para el trabajo narrativo, pero con el agregado de nuestro proceder lingüístico-semiótico: la tabla conjunta la terminología del área de prompts con algunas de la lingüística y la semiótica. La consonancia de estas tres áreas es otra forma de establecer la pertinencia de este método. Ya con el agregado nuestro, la tabla dispone el tema, el rema⁷⁶, un cotexto⁷⁷ para el *contexto*, el género del cuento y también cuestiones de estructura. El prompt seleccionado A no cuenta con un cotexto que actualice su contexto. La ausencia de este no es precisamente una falla

⁷⁶ Aportación de información nueva al tema de una oración, van Dijk 36.

⁷⁷ Elemento que aporta una actualización al contexto al ligar a un signo características distintas a las de su relación significativa “normal”, Eco, *Lector in fabula* 28-29.



siempre y cuando el contexto sea claro. En este caso, se está diciendo a la IAG que se busca escribir cuentos de ficción, simple y sencillamente eso. Las justificaciones para ello no son necesarias. El formato y el tono en que debe operar la IAG, que también se sugieren en el primer video sobre los bloques de prompts, vienen dados por la explicitación del género y por la lectura connotativa de los sememas “misterio” y “fantasmal”.

Tabla del prompt A

	Contexto	Tarea	Ejemplo
Género	Soy escritor de cuentos de ficción y quiero escribir un cuento del género policiaco.		[...del género policiaco]
Cotexto			
Tema		Proporcioname una lista de cinco ideas que desarrollen el tema de un asesinato...	
Rema		...misterioso.	
Estructura (actos)			Las condiciones para la trama son que primero los personajes que investigan el asesinato deben sospechar que el asesinato ha sido cometido por una entidad fantasmal cuando en realidad fue cometido por una persona.

Cuadro 4

Esta exhibición nos permite confirmar la pertinencia del prompt ante otras posibilidades. Por ejemplo, el caso B, como se muestra a continuación:

- B) Quiero convertirme en escritor de cuentos de ficción a través del uso de inteligencias artificiales generativas. Proporcioname una lista de cinco ideas para escribir un cuento interesante y creativo del género policiaco. El tema debe ser el de un asesinato misterioso. Las condiciones para la trama son que primero se debe sospechar que el asesinato ha sido cometido por una entidad fantasmal cuando en realidad fue cometido por una persona.



- a) Quiero convertirme en escritor de cuentos de ficción
- b) a través del uso de inteligencias artificiales generativas.
- c) Proporcioname una lista de cinco ideas para escribir un cuento interesante y creativo del género policiaco.
- d) El tema debe ser el de un asesinato misterioso.
- e) Las condiciones para la trama son que primero se debe sospechar que el asesinato ha sido cometido por una entidad fantasmal cuando en realidad fue cometido por una persona.

Tabla del prompt B			
	Contexto	Tarea	Ejemplo
Género (narrativo)	Quiero convertirme en escritor de cuentos de ficción...		
Cotexto	...a través del uso de inteligencias artificiales generativas	Proporcioname una lista de cinco ideas para escribir un cuento interesante y creativo...	
Género		...del género policiaco	[...del género policiaco]
Tema	[El tema debe ser el de un asesinato...]	El tema debe ser el de un asesinato...	
Rema		i) [...cuento interesante y creativo...] ii) ...misterioso.	
Estructura (actos)			Las condiciones para la trama son que primero se debe sospechar que el asesinato ha sido cometido por una entidad fantasmal cuando en realidad fue cometido por una persona.

Cuadro 5



El número de elementos permanece constante, mas se ha complicado la instrucción con un cotexto⁷⁸ y se ha añadido una ambigüedad excesiva al mensaje al solicitar del Transformador un cuento dentro de lo “interesante” y lo “creativo”. El cotexto estaría reforzando la relación usuario-IAG con una determinación de “uso de IAG como herramientas para la escritura”, información la cual podría presentar varios caminos disyuntivos no deseados, introducir nuevos *tokens* innecesarios al *embedding*⁷⁹ y aumentar las probabilidades de interferencias. Estos enunciados de los cotextos y remas innecesarios podrían incluso verse enunciados directamente en el discurso del cuento. Ahora, también el modelo A resultaría más satisfactorio en tanto que la tarea condensa también el tema, cuando en B podría ser tomado también como parte del contexto de la tarea. El prompt A economiza, no agrega actualizaciones innecesarias y apela a un contexto más claro que refuerza más efectivamente la tarea.

De cualquier modo, ambos modelos A y B son satisfactorios en tanto que tiene la intención de poner en marcha el *brainstorming* a través de la tarea (*task*) “proporcioname ideas”, con un verbo en imperativo y directo. El usuario podría luego hacer las correcciones pertinentes. Pero ya que buscamos un resultado óptimo, un resultado *modelo*, optamos por A.

A continuación de esto, cualquier exceso de ambigüedad debe ser cuidado, por lo que las mismas nociones con las que se construyó el prompt deben continuar activas durante el resto de la interacción. Por ejemplo, no proseguiríamos mediante el caso: “divide la historia en capítulos para hacerlo más ameno al lector...”. Ello podría generar un cambio radical en todo el cuento por el movimiento del *embedding* del Transformador. Bastaría con solicitar la división en capítulos.

⁷⁸ La selección que hemos hecho para este no está motivada; el punto es que cualquier actualización cotextual involucraría información innecesaria —para este caso de la creación narrativa; otros usos dependerán del contexto y finalidad— que podría intervenir en el proceso del Transformador.

⁷⁹ Véase capítulo 1.



De acuerdo con lo establecido para proceder en el nivel cooperativo beta, en el proceso de iteración se introdujo una instrucción más que corregía el ejemplo: en la etapa inicial del prompt se le indicó a ChatGPT4 indirectamente que los personajes investigarían el suceso problemático. Se tomó la decisión de añadir la creación del personaje de un detective, siendo esta la única disyunción importante por parte del usuario para el posible entramado que ChatGPT4 estuviera construyendo desde un inicio.

Para terminar, se proporciona, una vez concluida toda iteración, una instrucción final que pondrá en marcha la creación del cuento:

“Me gusta la idea trabajada X_i , y me parecen útiles los elementos $Y \dots Y_n, Z \dots Z_n$, etc. Escribe un cuento utilizando la idea X_i para la trama y que utilice los elementos Y_{i-n} y Z_{i-n} , etc. Debe contar con una introducción, un desarrollo, un clímax y una resolución. Divídalo en varios capítulos breves. La longitud final del cuento debe estar entre las N y M palabras.”

Por cuestiones de extensión, hemos decidido que el cuento se ubique entre las 1200 y 1500 palabras, espacio suficiente para que la IA ejecute lo más óptimamente posible la instrucción, adecuándose a todas las especificaciones expuestas, incluida aun la expectativa de la definición de cuento, por ejemplo. No obstante, ChatGPT4, por decisión propia (?) o por un fallo, ha emitido un cuento con una extensión menor a las 1200 palabras. Este resultado no invalida el experimento, al contrario: es de anotarse.

Así concluye una metodología que trabaje de la mano con ChatGPT4, con cualquier IA de red neuronal Transformadora, para la creación de un cuento modelo de muestra.



“¿Qué tan inteligente es una IA, Case?”

“Depende.

Algunas no son mucho más listas que los perros.

Mascotas. Igual cuestan un fortuna.

Las verdaderamente inteligentes son tan inteligentes como

la poli de Turing les permita serlo.”

Neuromante, William Gibson, 1984



Capítulo I. Autores y Lectores: las estrategias textuales de ChatGPT para la creación literaria

1.1 ChatGPT como autora empírica

¿Se puede tratar a una IAG como una autora empírica? La existencia de materiales que unifican en autoría a humanos e inteligencia artificiales se ha adelantado ya a esta interrogante. Pero lejos de simplemente aceptar la cuestión, existen algunas particularidades que repasar, que nos permitirán afirmar cómo es esta autoría empírica de manera precisa. De modo que nuestra respuesta a la pregunta es, de entrada, *sí, pero con marcadas diferencias respecto al autor empírico humano.*

En el marco teórico se ha definido el papel del autor empírico como una figura que trasciende su presencia en el texto a la manera de un estilo, una estrategia actancial, ideológica e incluso gramatical. ¿Qué corresponde a la IAG en este respecto? Por un lado, sería ocioso considerar a ChatGPT como una muestra de su época, a la manera de cómo se considera a un autor empírico humano del cual se puede aprender sobre su poética y su estilo que influyen en su manera de escribir. ChatGPT no es una “autora del siglo XXI, de X país y apegada a X movimiento literario”. Ya que no es posible construir un perfil de esta naturaleza, debemos buscar otra manera en que designar a los autores empíricos involucrados, porque es a partir de estos que se generará la estrategia textual denominada Lector Modelo, cuya definición es necesaria para cumplir con el objetivo de verificar la cooperación interpretativa con el texto generado.

En teoría, la IA debería trascenderse también en el discurso narrativo a través de sus diferentes componentes, como los nombrados para el autor empírico humano en la teoría de Eco. La cuestión es que el texto de la IAG tiene un tratamiento, un procesamiento y una génesis distinta que requiere



de este examen más a profundidad. De modo que para alcanzar una formulación precisa que la proponga como autora empírica, debemos rescatar la base de los fundamentos de una IA, de las IAG de Transformador en particular, y extraer algunas conclusiones que, de acuerdo con el concepto del autor empírico, se reflejen en la formulación de la estrategia textual que es el LM y en el AM. En otras palabras, para concluir satisfactoriamente sobre las IAG como autorías empíricas para la escritura creativa, lo haremos considerando sus *estrategias pretextuales que se podrían ver reflejadas luego en la creación textual*. Estas estrategias se traslucen por la forma de operar que tienen las IAG. Pero dado que los rostros tanto del autor empírico humano y de la IAG se difuminan en este ejercicio de escritura asistida, la verdadera enunciación de la autoría empírica solo estará completa cuando, a la aclaración de estas estrategias pretextuales, se sumen las conclusiones sobre el autor empírico humano *en función del proceso de creación asistida con la IA*. Este, argumentamos, es el verdadero panorama de una autoría empírica compartida entre humano e IAG.

El concepto de las IA en general adopta una postura que las define como tecnologías cruciales para el desarrollo del siglo XXI y que potencia su desarrollo con el objetivo de mantener bien en pie ese concepto: A) *igualar o superar las capacidades del ser humano* (IBM Technology, “AI vs Machine Learning”, *Youtube*, 0:40). Lo mismo enunciaba Elaine Rich varias décadas atrás en su frase que recuperamos en la apertura del planteamiento del problema. Pero esta definición en un principio adolece del mismo “tropiezo” que, en una época no muy remota, Alan Turing exhibió en su famoso artículo, donde no definía en amplitud a las IA ni definía —si bien no es algo que ni el mismo Turing se propuso hacer directamente— lo que es la inteligencia. La inteligencia de las máquinas, o la demostración de capacidades humanas es todavía un fundamento que se espera ver en realmente en estas tecnologías. El primer paso para dilucidar la figura de la autora



empírica máquina estaría en desmontar esta creencia. Si sus capacidades aspiran a la igualación o superación de las capacidades humanas, esperaríamos obras literarias que hicieran un eco diferente en la cultura, sin embargo, no es así; parece ser que las preocupaciones en este campo de la creatividad y la literatura son más por las posibilidades de lo que pueda suceder y no por una obra concreta que ya haya cambiado las cosas.

Si volvemos a Turing, encontraremos que él mismo se encargó de abordar cuidadosamente este aspecto de las máquinas como agentes pensantes y de prever las constantes fallas a las que esta aproximación a las máquinas estaría sometida:

¿No pueden las máquinas realizar algo que debería describirse como pensar, pero que es muy diferente de lo que hace un hombre? Esta objeción es muy fuerte, pero al menos podemos decir que si, no obstante, se puede construir una máquina para jugar satisfactoriamente el juego de la imitación, no debemos preocuparnos por esta objeción⁸⁰ (“Computing Machinery and Intelligence” 29).

Y, sin embargo, la aguda observación de Turing ha terminado por no conducir al descubrimiento de la inteligencia de las máquinas, sino a algo diferente⁸¹. El juego de la imitación es una visión reduccionista de la inteligencia humana, pese a que es un singular ejemplo conceptual, y por el cual otros experimentos se afianzaron bajo la misma determinación a probar o negar la posibilidad de que las máquinas pueden igualar esta inteligencia. De manera que, para muchas cuestiones, es una necesidad que el planteamiento de Turing deba ser observado desde otra perspectiva. Ya hemos dejado claro en apartados previos que existen consistentes faltas de

⁸⁰ En el original: *May not machines carry out something which ought to be described as thinking but which is very different from what a man does? This objection is a very strong one, but at least we can say that if, nevertheless, a machine can be constructed to play the imitation game satisfactorily, we need not be troubled by this objection.*

⁸¹ En 2.2.2 se encuentra la discusión exhaustiva al respecto, echando mano en especial de Searle y su planteamiento del juego del cuarto chino.



comprensión semántica y pragmática en las operaciones de estas tecnologías y sus antecesores, cuestiones que, en primera instancia, ya las separa mucho de la tan añorada inteligencia humana —o al menos de su capacidad cognitiva—, pero en esta ocasión no es suficiente permanecer en este señalamiento. Si no se habla de inteligencia, habría que definir, cuando menos, qué es aquello que las hace parecer que profesan de una inteligencia casi humana.

La definición a la que realmente apela Turing es una definición desde cierto punto de vista de la *capacidad performativa, de rendimiento*, que tienen las máquinas, en este caso de las computadoras, para llevar a cabo algo. De esta manera se expone en un comentario a su ensayo, en una edición más reciente del mismo:

Turing parte de una equivocación. Ahora sabemos que lo que va a considerar no es si las máquinas pueden pensar o no, sino si pueden hacer o no lo que los pensadores como nosotros podemos hacer —y, si es así, cómo—. Hacer es una capacidad de rendimiento, empíricamente observable. Pensar es un estado interno. Se correlaciona empíricamente observable como actividad neuronal (si tan solo supiéramos qué actividad neuronal corresponde al pensamiento) y su calidad asociada introspectivamente observable como nuestro propio estado mental cuando pensamos. La propuesta de Turing resultará no tener nada que ver ni con la observación de estados neuronales ni con la introspección de estados mentales, sino solo con la generación de una capacidad de rendimiento indistinguible de la de los pensadores como nosotros⁸² (Harnad, Comentario a “*Computing Machinery and Intelligence*”, 23).

⁸² En el original: *Turing starts on an equivocation. We know now that what he will go on to consider is not whether or not machines can think, but whether or not machines can do what thinkers like us can do – and if so, how. Doing is performance capacity, empirically observable. Thinking is an internal state. It correlates empirically observable as neural activity (if we only knew which neural activity corresponds to thinking!) and its associated quality introspectively observable as our own mental state when we are thinking. Turing's proposal will turn out to have nothing to do with either observing neural states or introspecting mental states, but only with generating performance capacity indistinguishable from that of thinkers like us.*



Turing todavía no hablaba en términos de inteligencia artificial *per se*, mucho menos de lo que conocemos ahora como Grandes Modelos de Lenguaje (LLM), así que no puede hacerse una lectura similar a las tecnologías que nos empiezan a ser tan familiares ya en tiempos actuales. Aun así, esta conceptualización de la *performance capacity*⁸³, carece todavía de sumar la ausencia de capacidades semánticas y pragmáticas y aclarar cómo se acercan las IA a eso que de alguna forma definimos por pensamiento, que a su vez acerca a pensar que son máquinas inteligentes. De momento es suficiente con entender que la propuesta de Turing comprende que las máquinas pueden exhibir un manejo a fondo de las capacidades para seguir las reglas de un juego, sobre una base de datos bien cimentada que le indique cómo llevar a cabo este juego. Unas capacidades, subrayemos, *óptimas*.

A raíz de esta observación, que nos parece válida en tanto que de entrada desviste la idea a veces incuestionable de “inteligencia” en las máquinas, y para ahondar en esas capacidades de operar de las IA, optamos por conceptualizar a las IA como A’) *Agentes racionales*. Según Stuart Russell y Peter Norvig en Artificial Intelligence: A Modern Approach, un agente es algo que actúa, lo cual permite un número de *capacidades performativas*, mientras que un agente racional además busca los mejores resultados, o *los mejores resultados esperados*, a su actuar (4). Esta definición es válida y actual conceptualmente y de hecho coincide con la idea de Harnad. Pero es apenas la primera pieza que conforma el rompecabezas que es hacer una definición completa.

Reunimos dos argumentos, A y A’. El último refina lo que quedaba de general en el primero. Lo que sigue para validar sobre todo esta última aproximación sobre la IA, es que esa capacidad de llevar a cabo los mejores resultados, y de igualar y/o mejorar las capacidades performativas de los seres humanos, debe estar, por supuesto, implícito en el funcionamiento de estas tecnologías.

⁸³ En lo que respecta al funcionamiento de las IA y las IAG, conservaremos algunos de los términos en inglés para preservar la precisión de estas explicaciones.



Ya sea por su imitación del cerebro humano, por su capacidad para igualar su lenguaje, comprenderlo y reproducirlo, etc. Para conceptualizar estas nociones, diremos, pues, que una de las bases de la IA es B) el *aprendizaje*. Este será, más adelante lo veremos, el fundamento de inteligencias artificiales generativas como ChatGPT4. Los mismos Russell y Norvig lo mencionan en su manual: “Todos los agentes pueden mejorar sus capacidades a través del aprendizaje”⁸⁴ (59); y es, a su vez, el mayor legado que el artículo de Alan Turing dejó al mundo de la IA: ese apartado final en su ensayo, titulado, precisamente, *Machine Learning*, nombre que ha adoptado una de las ramas cruciales en la ciencia de la IA.

El aprendizaje de las IAG —y en esto Turing conserva intacta su vigencia— es un proceso a la vez similar y a la vez diferente del de los seres humanos. Es, obviamente, un proceso mucho más sistemático. Habría que afianzar esta noción de aprendizaje.

Las inteligencias artificiales generativas, para este caso, ChatGPT, que entra dentro de los LLM, tienen su fundamento en el B') *aprendizaje con base al entrenamiento, refinamiento (finetunning) y la retroalimentación (feedback)*. La capacidad generativa de estas IA, como su nombre lo especifica, es posible solo a partir del preentrenamiento desde una base de datos (a grandes rasgos: texto tomado de internet⁸⁵), conocimiento que será codificado, traducido a números en grandes matrices dentro del cerebro de la IA que es la red neuronal, para luego aprender a predecir segmentos de texto y más tarde refinar ese aprendizaje y repetir el proceso. Esta es, a muy grandes rasgos, el proceso de una IA generativa, de un LLM. Ahora bien, en los detalles de

⁸⁴ En el original: *All agents can improve their performance through learning*.

⁸⁵ La pila, *The Pile*, para dar un ejemplo, fue un corpus textual en inglés de 825 GB, que se utilizó para pre entrenar grandes modelos de lenguaje. Se componía de 4 grandes categorías: texto académico (PubMed, ArxiV), textos de internet en general (Common Crawl, Wikipedia), prosa (datos de chats, diálogos de películas), y misceláneos (Github) (Jurafsky, Martin, Speech and Language Processing: An Introduction... 212). El preentrenamiento conlleva varias acciones, tales como la supervisión de este, la cual puede ser no humana o humana, y el filtrado de información, así como reconocimiento de los peligros de estas grandes cantidades de absorción de *data*.



este proceso se encuentran claves que recolectar, por lo que se vuelve necesario establecer otras definiciones y aclaraciones de su funcionamiento.

Una de las claves de la definición de un modelo de lenguaje grande⁸⁶, es el concepto de la *predicción* de texto. Siendo el prompt del usuario una cadena de palabras, por ejemplo, una pregunta que debe ser contestada, o una instrucción, en el funcionamiento de los LLM está el continuar esa cadena de palabras: con una respuesta, con un mensaje, con un texto. A eso se refiere la predicción de texto:

Un modelo de lenguaje es un modelo de aprendizaje automático que predice las siguientes palabras. De manera más formal, un modelo de lenguaje asigna una probabilidad a cada posible palabra siguiente o, equivalentemente, proporciona una distribución de probabilidad sobre las posibles palabras siguientes. Los modelos de lenguaje también pueden asignar una probabilidad a una oración completa⁸⁷ (Jurafsky, Martin, Speech and Language Processing: An Introduction... 32).

Y así, en ese cálculo probabilístico, el texto que aparezca será el que ha tenido las mayores probabilidades de aparecer. La probabilidad es calculada con base al contenido analizado del prompt del usuario para este caso.

⁸⁶ Nuestro trabajo se enfoca en las inteligencias artificiales del tipo LLM, como los chatbots, como ChatGPT4, Copilot, Gemini, etc. No todas las IA son LLM, pero todos los LLM son inteligencias artificiales, y los modelos nombrados tienen un “cerebro” definido como red neuronal. No todas las redes neuronales son para trabajos que se encargan de texto, también procesan imágenes y audio, por ejemplo. En nuestro caso, al utilizar ChatGPT4, estaremos utilizando como sinónimos la mayoría de las veces los conceptos de chatbot, LLM, IAG, Transformador y red neuronal, aun cuando cada concepto tenga su definición particular. Solo debe tomarse en cuenta que lo que liga a estos conceptos es la capacidad de usar el lenguaje escrito, de trabajar con él, generar textos, etc. La disciplina dentro del *Machine Learning*, que está a su vez dentro de la ciencia general de la IA, encargada de fortalecer esta función del lenguaje es comúnmente la NLP, la programación del lenguaje natural.

⁸⁷ Original: “A language model is a machine learning model that predicts upcoming words. More formally, a language model assigns a probability to each possible next word, or equivalently gives a probability distribution over possible next words. Language models can also assign a probability to an entire sentence”.



A todo esto le siguen diferentes fases para continuar refinando el texto, como el ya mencionado *finetunning* y la retroalimentación humana, directa o por aprendizaje, durante la conversación mantenida en el chat. De momento no ahondaremos en estos procesos, será en otros puntos de esta investigación donde se aclaren. Ahora bien, el proceso, sintetizado de esta manera, aparenta sencillez, pero lejos de eso, la cantidad de datos, prácticamente incommensurable, y de funciones ejecutadas simultáneamente tan solo para la predicción de la siguiente pieza de texto, o palabra, dotan de una complejidad matemática y operativa muy amplia y extensa a las IAG. Tan es así que una fuente define a los LLM como “una sofisticada función matemática que predice qué palabra viene a continuación para cualquier pieza de texto” (3Blue1Brown, “*Large Language Models explained briefly*” 0:37). Este tipo de ejecuciones tan complejas solo son posibles a través de tecnologías más refinadas y potentes, respaldadas, por supuesto, por el hardware adecuado, de las que son poseedoras IA más actuales como ChatGPT. Habría tan solo que considerar su funcionamiento, una buena parte del cual se halla en las siglas GPT: Transformador Generativo Preentrenado. Los dos últimos términos son claros en cuanto que corresponden con lo ya expuesto, la gran diferencia estriba en el Transformador: la red neuronal, el cerebro de ChatGPT y los LLM vigentes en el mercado actual.

En síntesis, diremos que la base del funcionamiento los LLM es la C) *predicción*. Pero el cómo la IA obtiene la certeza de que el texto que va a predecir y formular es válido, el cómo les asigna números a esas probabilidades, tiene que ver con el preentrenamiento. Principalmente de esas porciones de texto (y luego del *finetunning* y el *feedback*) obtendrá el conocimiento necesario que luego procesará, para poder, finalmente, entregar el texto adecuado en el *output*. Este proceso es iterativo, y dadas las cualidades de agentes racionales y su calidad de aprender, la idea es que el proceso sea cada vez más refinado y los resultados cada vez mejores.

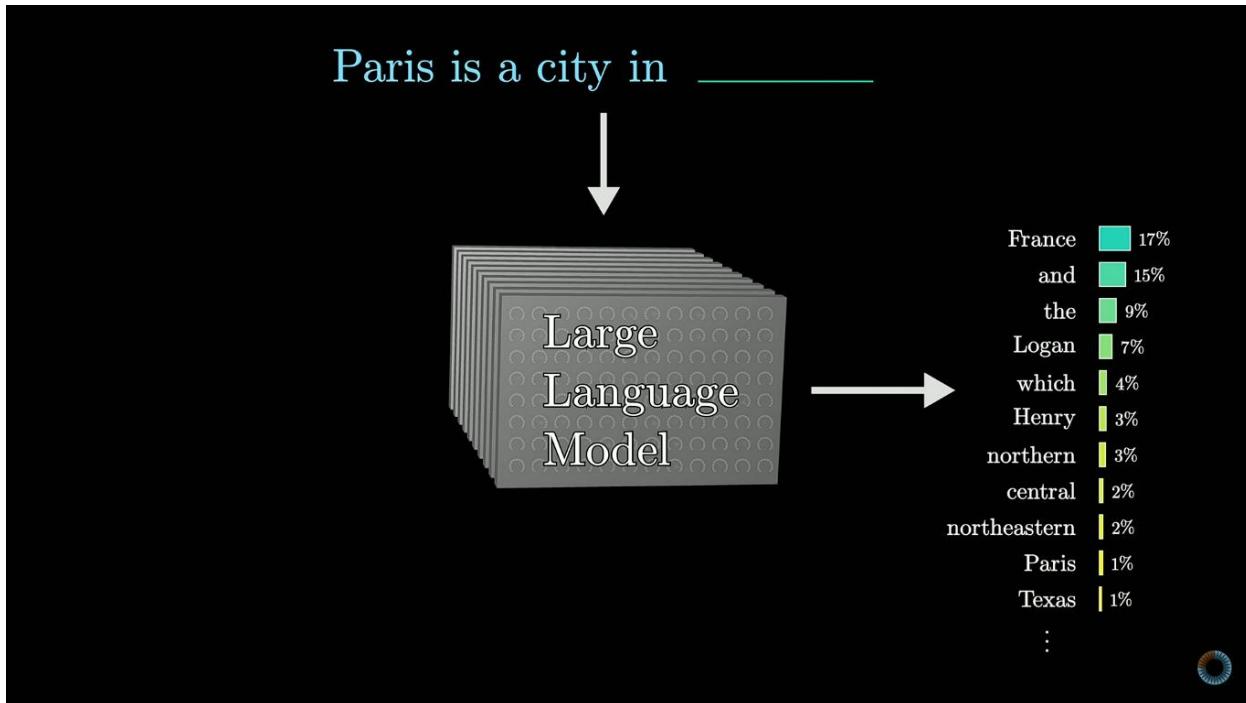


Figura 3. Fuente: 3Blue1Brown: “*Large Language Models explained briefly*”.

Pero para que un LLM sea un LLM, tiene que comprender esos textos de la fase de preentrenamiento lo suficiente como para saber de qué habla. En 2.2 entraremos en una larga discusión al respecto. De momento solo definiremos que este proceso de lectura de textos da paso a algo que se denomina como *embedding*, o incrustación, mecanismo que opera sobre la base de la semántica vectorial y sobre la selección probabilística de palabras, o *tokens*⁸⁸, y que está sometido a su vez a los procesos de *tunning* y *feedback* (3Blue1Brown, “*Transformers (how LLMs work) explained visually | DL5*” 13:36). El ejemplo de la figura 3 conceptualiza la naturaleza de esta predicción. Antes de que estas palabras aparezcan en fila con un valor porcentual, podemos

⁸⁸ Los LLM convierten el texto del preentrenamiento en texto normalizado a un sistema de lectura propio de estas IA. El texto se separa en unidades que se denominan como *tokens* (Jurafsky, Martin, Speech and Language Processing: An Introduction... 4-5). No se trata de una simple separación silábica ni por caracteres: la *tokenización* es un mecanismo propio de la programación del lenguaje natural y tiene sus propias reglas de segmentación. Asimismo, existen otros tipos de normalización de texto, como la lematización, el *stemming* o derivación y otras segmentaciones. En este apartado sólo es necesario reconocer que las palabras, cifras y demás caracteres del texto aprendido por los LLM serán considerados como tokens dispuestos en el espacio vectorial.



imaginarlas dispuestas en un campo semántico a manera de vectores. Lo que determina su posición en el campo es un valor arbitrario que luego se va modificando. Digamos que uno de los cuadrantes del campo vectorial refiere a nombres del reino animal. La red neuronal irá configurando palabras para que todas las que tengan un valor entre 5 y 10 sean palabras que, claramente, pertenezcan al cuadrante del reino animal. Al parámetro, o valor, que se le da a los *tokens* en este proceso, se le conoce prioritariamente como *weight*⁸⁹. Pero estas palabras, o *tokens*, no están aisladas en el espacio con su valor numérico: también tienen relevancia por la proximidad unas con otras, a lo cual se le denomina, dentro de la red neuronal Transformador, atención (*Attention*), algo así como una “capacidad de atención de sí misma” que tendría la red neuronal, que le permite prestar atención simultánea a los conceptos que alberga, relacionarlos contextual y matemáticamente unos con otros:

El Transformador es una red neuronal con una estructura específica que incluye un mecanismo llamado autoatención o atención multicabezal. La atención puede concebirse como una forma de construir representaciones contextuales del significado de un token prestando atención e integrando la información de los tokens circundantes, lo que ayuda al modelo a aprender cómo se relacionan los tokens entre sí a lo largo de amplios períodos⁹⁰ (Jurafsky, Martin, Speech and Language Processing: An Introduction... 184).

⁸⁹ El concepto de *weight* es bastante amplio y abstracto en veces, pero para fines de este ejercicio es suficiente recordar que a un *token* se le asigna un valor de *weight*, y que esta asignación es un resultado del reiterado proceso de aprendizaje del LLM, que empieza, como se ha señalado, desde el preentrenamiento. Otro ejemplo conceptual es el de asignar a la palabra “increíble” un valor de *weight* positivo y a la palabra “abismal” uno negativo (Jurafsky, Martin, Speech and Language Processing: An Introduction... 79). Es una asignación arbitraria que instaura parámetros para que palabras relacionadas con increíble tengan un valor positivo y las relacionadas con abismal uno negativo, valores que luego irán modificándose. Si hacemos hincapié en esta explicación es por la complejidad del concepto, y porque luego existen otras maneras de determinar el valor específico de un *token*, como con el concepto de *bias*, que no debe confundirse con los sesgos semánticos (ver 3.2): tanto el *bias* como el *weight* van alterando los valores de los *tokens*, constantemente refinando la comprensión de estos y representan, en suma, el esfuerzo del iterativo proceso de aprendizaje que ejecuta la red neuronal, lo cual es a su vez la forma de crear significado que tiene el LLM.

⁹⁰ Original: “The transformer is a neural network with a specific structure that includes a mechanism called self-attention or multi-head attention. Attention can be thought of as a way to build contextual representations of a token’s



Todo este largo proceso de incrustación (*embedding*) y asignación de valores (*weights*) y la asimilación de estos (*attention*) es, digamos, el proceso de conformar una enciclopedia interna por parte del LLM. Enciclopedia a la que accederá para verificar valores, conceptos y después seleccionar los *tokens* más probables con base a la entrada de texto, o prompt, que ha proporcionado el usuario. Si atendemos al ejemplo de la figura 3, antes de que “Francia” aparezca como la continuación más probable al enunciado de “Paris es una ciudad de...”, el LLM ya debe haber tenido bien codificado en su red neuronal los *tokens* “Francia”, “Ciudad”, “París”, “es”, “de”, etc., con relaciones establecidas entre ellos, *weights* asignados y el proceso de *attention*. Dicho otro modo, este proceso es lo que hace que a los LLM obtener el significado de un texto⁹¹.

Todo está dado por los números en los Transformadores (3Blue1Brown, “*Large Language Models explained briefly*” 4:58). Las entradas, incluso las imágenes, son traducidas a números, a matrices numéricas. La matriz numérica que enmarca a los vectores cuyas cifras definen los *tokens* es la forma aproximada al *significado* con que cuentan los Transformadores. Todo este caos numérico evidencia que estas redes neuronales operan bajo el criterio del *determinismo* de que la misma acción repetida un número constante de veces, de que la misma entrada otorgará, idealmente, el mismo resultado siempre, aun con la cantidad quasi infinita de datos y procesos que están en juego —a no ser que los valores de las entradas y los *weights* se modifiquen significativamente en el proceso de aprendizaje, o *finetunning*—. De manera que, al observar todo este proceso complejo, la conclusión que hay que sostener es la de este determinismo numérico es la base conceptual de las inteligencias artificiales actuales. Casi una poética, podríamos decir desde la literatura: una manera de regir su funcionamiento, porque es precisamente la única manera de

meaning by attending to and integrating information from surrounding tokens, helping the model learn how tokens relate to each other over large spans”.

⁹¹ En 2.2 se discute la validez esta noción, que de momento es válida enunciar para la formular la imagen del autor empírico.



entender las cosas, el mundo: una epistemología basada en los números, en la *data*⁹², hecho que desemboca en otras partes que hemos destacado de su funcionamiento, como la cualidad de agentes racionales, por ejemplo. Los números determinarán la meta de lo que es óptimo y que debe ser alcanzado. Bajo este criterio, cualquier tropiezo del determinismo numérico es, simplemente, un error (uno que no debería desestimarse del todo⁹³).

En síntesis, *una IAG como ChatGPT4 es una autora empírica según sus capacidades para demostrarse como A') agente racional, es decir, que buscará el mejor resultado esperado según la entrada proporcionada por el coautor humano. Este criterio que busca alcanzar va precedido por todo el bagaje literario de la IAG, y el tratamiento que este recibe por su funcionamiento, es decir, a través de su proceso de B') aprendizaje, que va de su alimentación desde bases de datos, a la configuración de su epistemología con el embedding vectorial, el tunning y el feedback, y que tiene como resultado una C) predicción textual considerada como el resultado óptimo de una selección. Estos elementos pretextuales han de determinar sus estrategias textuales visibles en su conformación del Autor Modelo.*

Hay mucho que entra en juego en estas estrategias cuando las confrontamos con el acto de crear literatura. Tan sólo basta considerar la fase de preentrenamiento de la IAG para imaginar la cantidad de contenido a la que es sometida. En estas bases de datos aprenderá, por ejemplo, qué es un cuento, qué un género literario, y los códigos lingüísticos y semióticos y narrativos más básicos en lo que respecta a la creación de una historia: códigos gramaticales, códigos de los actantes, de las tramas, etc. Pero toda esta *data* será luego sometida a tratamiento. ¿Cómo entenderá qué es la literatura? ¿De qué parte de internet determinará la mejor definición para lo que es un cuento? ¿De

⁹² Véase en 3.4 el tratamiento que desde las Humanidades Digitales se propone con base a la epistemología, y aun ontología, que acompañan al concepto de *data*. Asimismo, en 2.3 se vuelve a poner en cuestión la pertinencia de este determinismo numérico desde la perspectiva del arte y la estética.

⁹³ Véanse capítulos 2.3 y 3.2: para la cuestión del *unheimlich* y las alucinaciones.



dónde obtendrá las características con las que definirá a los personajes, si el coautor humano no le da un perfil completo? ¿De dónde viene la información con la que llenará los huecos en la instrucción del coautor humano? Aunque lográramos responder a estas preguntas, su forma de comprender esta información, si bien tiene ya una base constante y sólida, siempre habrá de estar en un constante refinamiento. Por todo esto, no debemos dejar de subrayar la importancia del tratamiento que estamos haciendo aquí: estamos buscando formular las categorías de autor empírico, Autor Modelo, etc., pero cuando una de las autorías es una IA, hay otras implicaciones que necesitamos conocer para alcanzar la formulación adecuada de esas categorías y después alcanzar otras más, como las de obra abierta y obra cerrada. De modo que estas indagaciones de lo que hemos definido como estrategias pretextuales de una IA nos parecen las mínimas a considerar. El conocer que son agentes racionales, que son preentrenadas de enormes bases de datos, que la base de su funcionamiento textual es la predicción de palabras respecto a un prompt o texto y que estos *outputs* son luego refinados, representa la aproximación que podemos tomar desde la teoría literaria, desde la semiótica textual, para verificar qué está en juego durante la creación de textos. Sin embargo, tampoco hay que perder de vista que el trabajo llevado aquí, tanto por las restricciones de extensión, como por las dificultades que implica el conocimiento profundo de estas áreas ajenas a la teoría literaria, no es exhaustivo. La forma en que desde la literatura y la semiótica textual tenemos para subsanar cualquier hueco emergente ante esta labor, es completando el concepto de autor empírico, o sea, la otra mitad del trabajo: definir el papel del usuario humano como autor empírico, y entonces, conjuntar estas dos autorías empíricas y llegar a una síntesis de este juego colaborativo.



1.2 El papel del autor humano (empírico)

Este punto es en esencia más sencillo que el de la IAG, pero sólo si definimos en primera instancia cuál va a ser la participación específica del usuario.

La clasificación del coautor humano en el apartado metodológico nos llevó a considerar su intervención en cinco niveles. Esta clasificación podría abrir a otro tipo de observaciones en cuanto a creatividad e influencia del estilo del autor humano. No obstante, para términos de esta investigación, nos limitaremos a que la clasificación por niveles nos permitirá definir al autor empírico humano y, más adelante, a los AM y LM.

Dicho esto, pues, el nivel beta al que nos remitimos involucraría la creación de un cuento a partir de unos cuantos prompts dados por el usuario a ChatGPT. Mientras que un nivel alfa indicaría una creación donde la autonomía creativa corresponda casi por entero a la IAG, en el nivel beta la participación del usuario conlleva la toma de algunas decisiones principalmente en lo que a la disyunción de la historia atañe⁹⁴. Está luego, por supuesto, el *brainstorming*. ¿Qué tanto se involucra el usuario en la creación de ese espacio de ideas con las que trabajar, en la disyunción de decisiones creativas? En un nivel alfa, el *brainstorming* o no existe o es mínimo, sólo un planteamiento inicial, lo cual, desde una primera observación, no es lo ideal. Para que ChatGPT4 tenga un desarrollo provechoso debe existir esta iteración en todo el proceso que la haga, en una misma conversación, atravesar de nuevo los procedimientos en que involucra su funcionamiento basado en aprendizaje.

Esta consideración abre varias posibilidades, las primeras de las cuales son estéticas. Por ahora, basta decir que con la clasificación en cinco niveles estamos hablando de la intervención

⁹⁴ Esto fue lo que hicimos y se puede apreciar en el apéndice B al agregar ese “extra” a la historia, solicitando a la IAG la creación de un personaje más: un detective.



del usuario en términos cuantitativos. Lo cualitativo quedaría definido por el manejo de la estructura narrativa y aun por las correcciones estilísticas y prosísticas puntuales. De cualquier modo, un incremento en la intervención cualitativa en el texto debería ser proporcional a la intervención cuantitativa, a no ser que se trate de correcciones morfológicas apenas perceptibles, como en términos de sintaxis, o de algunas decisiones léxicas. Lo que sugerimos más ampliamente en 2.3 es que el usuario tome una participación más activa en el texto, en un nivel delta, digamos, o epsilon, a la manera de la escritora japonesa. Pero de ser así, surgiría entonces una duda ante este posible escenario: con niveles de intervención de la máquina tan bajos, ¿por qué un usuario necesitaría de una IAG para la creación de un cuento? Si el usuario tomara el control creativo en mayor parte, una posible respuesta a la pregunta es que se estaría utilizando a la IAG como una herramienta de soporte creativo para secciones específicas, y no como una herramienta *para escribir narrativa*. Pero esta última es la opción que nos interesa. Y este proceder empieza ya a formular al autor empírico humano: *todo empieza por su nivel de intervención*. Desde la comprensión de este fenómeno fue que sugerimos en el marco metodológico la posibilidad de mejorar la forma de cuantificar esta intervención. A un mayor entendimiento de esta, más diáfano se presentará el proceso de composición de textos mediante las IAG.

Desde este hallazgo definiremos al autor humano empírico. El autor empírico humano interviene en el texto, entonces, a manera de instrucciones y decisiones emitidas a la IAG. Estas instrucciones se determinan en términos principalmente de lo que será el tema del cuento, o la novela o el poema, y las disyunciones de la trama, disyunciones que se ven ajustadas por los tipos de personajes que el usuario pueda decidir, el narrador, etc. Esto es válido en un nivel general, o sea, en cualquiera de los niveles propuestos de IU.



Hemos mencionado instrucciones, decisiones, personajes, temas, etc., como parte de un conjunto de estrategias con las que interviene el autor empírico humano. Pero la realidad es que todo intento por rastrear los elementos específicos de la intervención del autor empírico humano resultaría en resultados posiblemente inexactos. Por ello, de momento una afirmación certera, que luego habrá de conducirnos a un mejor rastreo textual, es la de que el autor empírico humano, en este doble juego, *se define por su nivel de interacción con la máquina*, por la influencia con que intercede en todo el bagaje de la IAG, y que lo pone en marcha. Los elementos específicos podrían ser muchos y muy distintos, pero si los observamos desde la perspectiva de que todos implican aumento o disminución del nivel del usuario, estaríamos puntualizando mejor su calidad de autor empírico, evitando así caer en el peligro de creer que se puede separar perfectamente lo que hace el usuario de lo que hace la máquina. Esto sería posible únicamente si el usuario fuera lo suficientemente transparente con los movimientos que ha realizado y con el nivel de intervención que tuvo en el texto final⁹⁵.

De modo que, si unimos ambas observaciones, la del autor empírico ChatGPT4 y el autor empírico humano, el *autor empírico general* de las creaciones por medio de IA es *una instrucción*. Es, básicamente, *el prompt*. Ahora bien, el prompt y su ejecución no son sino la *sumatoria del nivel de intervención del autor empírico humano, con sus competencias implicadas para la toma de decisiones creativas al momento de formular sus instrucciones, y de las estrategias pretextuales de la autora empírica IA*. Y debido a que la IA es un agente racional, se puede esperar que cumpla

⁹⁵ Véase en 3.3 el concepto de transparencia introducido en el marco ético.



lo mejor posible con la instrucción del humano. Esto nos conduce a formular que hay un mayor peso en el *coautor humano* por una simple razón: *a mejor construida y más clara la entrada, mejor será el resultado en general entregado por la IA*. Ahora, tenemos otra forma de comprobar que el *autor empírico general* es esta sumatoria de intervención y competencias de ambas partes, si pensamos que la satisfacción de cierto nivel de calidad estética del texto generado depende también de la retroalimentación *entre máquina y usuario* y en la iteración de este proceso. No se pierda de vista que el coautor humano tiene el peso y autoridad para también decidir la longitud y lo intrincado de este proceso de curaduría, es decir, que mucho de ello deviene por su nivel y calidad de intervención. A todo esto, hay que recordar que, para lograr satisfacer los ideales estéticos del coautor empírico humano a través de la IAG, lo que importa, como dijimos antes, no es tanto el conocimiento literario como sí el conocimiento de una buena generación de prompts, que tenga en cuenta y conozca bien el funcionamiento de la IAG, de ahí que por un momento resulte fácil, si bien excesivamente retórico, decir que el autor empírico es el prompt.

Con base a estas conclusiones podremos formular ahora el Autor Modelo del cuento creado, el Lector Modelo de ChatGPT, y determinar la calidad de la cooperación interpretativa entre esta IAG y el usuario (obra cerrada u obra abierta). Porque ahora sabemos que el autor empírico general es el que generará, en el prompt y en el proceso de creación, implícitamente, una idea de Lector Modelo, que generará luego un Autor Modelo.

1.3 El Autor Modelo de “El Ritual de Medianoche”

Hay una aclaración que hacer respecto al título del apartado. Este se refiere al Autor Modelo *del cuento generado*. El apartado 1.4, y el título de esta tesis, en cambio, hace referencia al Lector Modelo *de ChatGPT*, implicando que es posible decir algo general sobre esta estrategia de



escritura de las IAG de Transformador. En efecto: es el resultado al que se espera llegar en dicho apartado. Con el Autor Modelo debería suceder lo mismo, no obstante, hay dos razones por las que este apartado no pone su peso en ese enfoque general: la primera es una complicación. El lector empírico es quien se genera una idea propia del Autor Modelo como una estrategia textual latente en el texto, y aunque lo hace con base a estructuras bien definidas y observables (actantes, temas, isotopías, macroproposiciones, estructuras de mundos posibles, estructuras ideológicas, inferencias, etc.), no deja de ser demasiado complicado generar una imagen clara, definida y controlada del lector empírico. Sólo podemos estudiar la presencia de estas estructuras como lectores empíricos con conocimientos de la teoría literaria y la semiótica textual. Luego, la decisión de mantener el enfoque del AM en el cuento generado responde a que la formulación del AM no estará del todo completa hasta que se hayan revisado los apartados 2.1 y 2.2 (del mismo modo que para definir del todo el Lector Modelo de ChatGPT habrá que revisar el apartado 2.3, pero como se verá en 1.4, es posible alcanzar todavía más conclusiones respecto al Lector Modelo después de todo lo reflexionado en los apartados que le preceden). De todas maneras, no debe perderse de vista la idea de que las conclusiones que intentamos alcanzar son más abarcadoras que solo su verificación en un cuento: porque estamos buscando delinejar la cooperación interpretativa no en un cuento, sino *en las inteligencias artificiales generativas*.

La búsqueda de un AM en el cuento realizado por ChatGPT4 no es otra cosa que establecer una hipótesis sobre la estrategia textual que se ha plasmado en el cuento (las estructuras). En particular, el AM se trata de una interpretación que se permite seguir el lector para introducirse en la cooperación con el texto, seguirlo en su propuesta⁹⁶. Para esto, como lectores empíricos,

⁹⁶ Como también se mencionó en el MT, para Luz Aurora Pimentel, el narrador es la figura clave en la formulación de lo que aquí definimos como Autor Modelo. Las selecciones de información del narrador, en términos cuantitativos y cualitativos, determinarían mucho de la forma de esta estrategia de lectura dirigida al lector. Incluso, menciona, remitirían en cierta manera también al autor empírico, dado que algunos narradores parecen establecer un perfil no



habremos de hacer las observaciones lingüísticas y semióticas que quepa hacer para definirlo: reparar en un uso especial de las palabras, en los actantes, en el o los temas propuestos, en el final de la obra, etcétera, hasta donde el texto nos lo permita. Pero no sólo se trata de construir una “encyclopedia” sobre el texto, que nos revele estas cualidades lingüísticas y semióticas. Las circunstancias de enunciación del texto también son importantes (Eco, *Lector in fabula* 91, 95) y la consideraremos, precisamente porque esas circunstancias las tenemos ya dadas por el establecimiento del autor empírico general. No obstante, sabemos que no todo lector empírico podría acceder a esta información sobre la naturaleza concreta de su autor empírico. En todo caso, podría tener solo una idea relativa a lo que es una IAG como ChatGPT4, y algo sobre sus capacidades. Para este ejercicio sí consideraremos lo que hemos concluido en el apartado anterior, pero debe tenerse en cuenta que para el proceso de interpretación, un lector que no formara parte de la investigación podría formular otro tipo de connotaciones sobre el texto al enterarse que ha sido escrito mediante ChatGPT, una IAG, si bien no muy alejadas de algunas de las conclusiones establecidas que le remitan sobre todo a pensar en lo formulado en los incisos A y A' del apartado 1.1. Por otro lado, recordemos que toda esta información es concreta de la semiótica textual y es útil sólo para el investigador. Un lector sin pretensiones investigativas no debería preocuparse por si alcanza el nivel de Lector Modelo, etc., y eso no interferirá, por supuesto, en su proceso de interpretación.

Sobre la base de estas consideraciones, analizaremos ahora el texto para hacer esta interpretación sobre el AM. Nótese que el configurar esta idea es ya parte de un proceso de interpretar el texto, pero, por supuesto, el enfoque será sobre las decisiones que el texto, su AM, ha tomado, y no todavía sobre la cooperación interpretativa como tal.

muy distante del que se conoce de los autores empíricos, históricos (174). Dada esta idea fructífera, en los apartados siguientes habremos de volver sobre la figura del narrador de la historia.



El capítulo I de “El Ritual de Medianoche” nos presenta a sus personajes estableciendo relaciones entre ellos y sobre ellos mismos: un grupo de amigos adultos de diversas profesiones. El AM llama a aquel encuentro un “juego”, según el subtítulo del capítulo (esta es otra estrategia textual: la de subtitular los capítulos), y un “experimento”, que es llevado a cabo por Eva, la psicóloga, quien desinteresadamente no informa a sus amigos de la verdadera naturaleza del evento. Esto informa el AM a su lector desde el inicio. Esto es ya una estrategia de formulación de una premisa, o de un primer acto. Las selecciones de los nombres no parecen remitir a significados mayores. El Lector Modelo estará figurándose ya un tema⁹⁷ para el cuento en este primer capítulo.

En los capítulos siguientes se sucede la muerte de María, y como parte de su estrategia, el AM deja entrever la posibilidad, para su LM, de que piense que el hecho estuvo en manos de un fantasma o ser espectral. Expresa esto de dos maneras: a través del juego de la ouija y de la generación de eventos espontáneos: “una ráfaga de aire frío recorrió la sala”; “las luces se apagaron por completo”, etc.; y con el respaldo del personaje de Ana, quien en el capítulo 2, está “convencida de que habían despertado a un espíritu maligno⁹⁸”. Según la estrategia que es el AM, Ana estaría intentando guiar la atención y credulidad del Lector Modelo: el AM del cuento quiere que se distraiga pensando esto, y si el lector, tal como establece un mago su truco de magia, ha aceptado el pacto de la “suspensión de la incredulidad”, lo hará. Sin embargo, hay varias marcas evidentes en el texto que, como veremos más adelante, intervienen con esta suspensión de la incredulidad, pero está claro que no está en la estrategia del AM que esto ocurra, *sino que es el resultado directo*

⁹⁷ El tema es la estructura principal de un cuento: “... el *topic* fija los límites de un texto” (129), y eso lo hace uno de los primeros movimientos interpretativos del lector, que puede proceder a formular desde el título del texto y de las circunstancias de enunciación (ejemplo: sabe que es un cuento que puede estar leyendo en una antología).

⁹⁸ Pimentel sugiere que hay frases en los textos que “tienen una clara función explicativa y orientadora para el lector” (168). Y esto está dictaminado, por supuesto, en lo que ella denomina programas de lectura inscritos en el texto narrativo, es decir, lo que nosotros denominamos Autor Modelo. Eventualmente estas indicaciones van dirigidas al Lector Modelo.



de las capacidades del autor empírico general, idea ya mencionada y que aparecerá a lo largo de otros de los ensayos de esta investigación.

El siguiente movimiento del AM es añadir la presencia de un detective para dar avance a la trama —como investigadores, sabemos que esta fue una decisión tomada por el coautor empírico humano—. El AM recurre a una primera elipsis donde, durante la cual los personajes iniciales contactaron y contrataron a este detective o este llegó como parte de todo el cuerpo de peritaje. Lo abrupto de esta elipsis y toda la información que consume, se puede argumentar, hiere la verosimilitud y ritmo del relato. El LM asumiría la imposibilidad de los personajes de resolver el conflicto por ellos mismos y deberá aceptar la repentina presencia del detective.

En este punto, el lector deberá convertirse, según lo desea el AM, en el mismo detective, y recolectar junto con él un número de pistas posibles que ayuden a una tirada de interpretación a la pregunta: “*¿qué o quién mató a María y por qué?*”. Este movimiento, el plantear esta pregunta está íntimamente emparentado con lo que se espera de un Lector Modelo. El objetivo del AM aquí es uno solo: llevar a su LM a la resolución del misterio; si cabe, antes que el personaje del detective. Su estrategia es no lineal: deberá disfrazar las evidencias como actos de un ente fantasmal. Esta decisión fue tomada por el autor empírico humano, y la calidad de esta estructura narrativa, el qué tanto puede el AM hacerle creer a su LM que, en efecto, un fantasma mató a la víctima, reposa por entero en el autor empírico general.

Al final del cuento, el AM activará dos cuadros intertextuales⁹⁹. El primero, con la resolución del caso. El cuadro típico, o estereotipado del género policiaco le dice al lector que el

⁹⁹ Los cuadros intertextuales según Eco, “son esquemas retóricos o narrativos que forman parte de un repertorio seleccionado y restringido de conocimientos” (120). De este modo, diferentes lectores tendrían diferentes tipos de experiencias. Quienes logren reconocer este tipo de cuadros en este texto, podrán prever mejor los acatamientos o violaciones a las reglas genéricas de los textos involucrados, mientras que aquellos que no disponen de los suficientes cuadros intertextuales, “están expuestas a gozar o a sufrir, por las sorpresas o soluciones espectaculares, mientras que el lector sofisticado las considera bastante triviales” (120).



detective *siempre* atrapará al malo, le cueste lo que le cueste, al menos en este enfoque decidido a *no tomar riesgos*. Y es así como sucede en el cuento de la IAG. Por otro lado, se activa, en el último párrafo, el cuadro intertextual que accede al género de la fábula: “la verdad siempre encuentra su camino” agrega un toque moralizante al cuento, esperando que el LM aprenda esta lección. Es de notar que en ningún momento en el prompt se aclaró que el conflicto debería resolverse positivamente para el detective. Esto, por supuesto, pudo haberse aclarado durante el *brainstorming*. No obstante, esta libertad otorgada deliberadamente a ChatGPT4 nos permite ver su funcionamiento en acción, vislumbrar su, digamos, poética respecto a cómo debería terminar un relato policiaco.

Como sugiere Luz Aurora Pimentel desde su postulado de los programas narrativos, otra forma de determinar cómo el texto quiere ser leído (AM) es a través de la estructura narrativa y discursiva que se conforma con la selección del narrador del relato. A través del enfoque que se ofrece del narrador, el lector tendrá su contacto con el mundo diegético del texto. Para este caso, sin ninguna intervención especial del usuario con el prompt, ChatGPT4 ha decidido optar por lo que típicamente se conoce como un narrador omnisciente, o, en términos de la teórica, el relato estaría enmarcado en la perspectiva del narrador, que es de tipo extradiegético con focalización cero, o libre. El Autor Modelo que opera en el texto “ha establecido” que la utilización de este tipo de narración es la estrategia, quizá, *ideal* para presentar al lector esta historia, descartando así una narración en primera persona o una narración que alterne su focalización entre personajes.

Tras estas observaciones, aventuramos que el lector empírico interpretaría a su Autor Modelo¹⁰⁰ del siguiente modo: el AM de este texto, tomando en consideración, por supuesto, que

¹⁰⁰ Considerando que para un lector no iniciado en la teoría literaria y en la semiótica textual, lo que aquí desvelamos como Autor Modelo será para él una visión más general sobre la trama o el texto en sí mismo, cuando no una confusión con su autor empírico, como se verá más adelante.



fue coescrito con ChatGPT4, recibe a su lector con una intención de generar expectativa: este tendrá una expectativa por ver cómo demuestra sus alcances creativos esta tecnología innovadora. Es una de las primeras connotaciones que surgirían al percatarse un lector de que el texto que se dispone a leer está coescrito con una inteligencia artificial. “¿Qué aporta esta IA, con su innovación tecnológica, al texto?” es, por tanto, una de las preguntas iniciales. Supongamos que el lector haya llegado a este cuento atraído por esa curiosidad ante las posibilidades de una IAG en la escritura creativa: esperará, cuando menos, un AM astuto en la construcción de su trama. Lo que se encuentra entonces el lector empírico, lo que construye con ayuda del texto, es *un AM que sencillamente se ha limitado a respetar muy cuidadosamente las reglas del género policiaco y del género narrativo*, según un lineamiento poético aristotélico. Y que, además, a manera de una fábula, *le ha dado la oportunidad de moralizarse en el conflicto de la verdad dentro del engaño*. Por otro lado, el léxico del AM no es desbordante y se centra poco en descripciones para favorecer el avance narrativo del conflicto del cuento. Es un AM que privilegia la presencia de un misterio que debe ser resuelto, entrega una pista falsa, y al final, una solución, y para esto usa como eje principal un actor capaz de resolver misterios, que encuentra, realmente, pocas dificultades a su paso. Quizá esto podría interpretarse como una necesidad por llegar pronto a la reflexión moralizante, y argumentar que es lo que tiene más peso, para que así el lector entrevea que la verdad siempre saldrá a la luz en cualquier situación, pese a que “... la oscuridad y el miedo a veces parecen invencibles”. Si esta es la conclusión del detective, y el AM ha querido hacer de su LM el detective, el lector deberá aceptar esta nueva convicción.

El investigador, como lector empírico y Lector Modelo de este cuento, que tiene la ventaja de haber delineado al autor empírico que se encuentra tras de la escritura, puede agregar lo siguiente: *el AM del cuento se configura en la unión de las propuestas de ChatGPT4 y de las del*



usuario humano, teniendo el mayor peso del control creativo la IAG. Este punto de partida de propuestas ha llevado a crear una estrategia que consiste en seguir una trama propia del género policíaco, y de establecer un tema (asesinato) y un argumento para el cuento que el lector puede enunciar a manera de macroproposición: un cuento sobre un crimen misterioso y su resolución. El AM se ha guiado por las marcas más relevantes de la instrucción, o prompt, del autor humano, para generar el texto de manera óptima: *de ahí que siga las reglas del género al pie de la letra, sin arriesgarse demasiado en las disyunciones de la trama.* Por esto antes hemos dicho que no se toman demasiados riesgos en la trama del cuento: no se puede porque no es lo que se le ordenó a la IAG. Esta hizo su trabajo perfectamente al generar algo apegado a lo que conoce como “cuento del género policíaco”. En todo caso, sabemos, para modificar esto el autor empírico humano debió haber tenido más intervención en el proceso creativo.

En síntesis, *el AM del cuento es una estrategia textual construida en mayor parte por la influencia de la alta cantidad de participación de la IAG, que sigue al pie de la letra las reglas del género policíaco, a partir de las instrucciones de su coautor empírico, ejecutadas de la manera óptima según le ha permitido su capacidad operativa.* En términos de las estructuras narrativas y discursivas, se puede decir que *el AM, a través de un lineamiento poético-aristotélico, intenta desviar débilmente las inferencias del lector antes de la anagnórisis, privilegia la resolución del misterio antes que otros elementos de la trama, y entrega una reflexión moral al lector.*

Estas conclusiones encierran perfectamente tanto las posibilidades de la doble autoría del cuento, involucrando así las capacidades de ChatGPT4, sobre todo como un agente racional, como la calidad del AM del texto propiamente dicho. Pero quizá lo más interesante de anotar es la reflexión general, que antes hemos dicho que no ha de quedar completa sino hasta revisar el apartado 2.2. Esta reflexión es en torno a que, sea cual sea el nivel de intervención del usuario, la



IA, precisamente como agente racional que es, y por sus cualidades implícitas, siempre se apegará estrictamente al prompt del usuario, ejecutándolo de la *mejor manera posible y al pie de la letra*.

1.4 El Lector Modelo de ChatGPT

Este apartado se centra sobre todo en la reflexión extendida sobre el Lector Modelo que proponen las IAG de Transformador. Podemos concluir de esta forma general en cuanto que ahora sabemos que mucho de los resultados dependen de factores antes establecidos, en especial, de la intervención del usuario. La última idea del apartado anterior, la que deduce en buena medida el AM de ChatGPT4, deberá también tomarse en cuenta para este apartado del LM. Sin embargo, las conclusiones deben ser más sustanciales, en tanto que, como ya se detalló en 1.3 y en el MT, generar una hipótesis de Lector Modelo es una labor más clara que la de “detectar” o interpretar una estrategia de Autor Modelo, porque la calidad de esta depende de la interpretación del lector empírico, figura extensa e indeterminada, mientras que aquella es más fácil de verificar¹⁰¹, en tanto que está en la previsión del autor empírico, y en este caso la ventaja es doble: como coautor empírico y operador de la IAG, que se ha preocupado antes por definirla en sus características operativas y estrategias pretextuales, el investigador tiene acceso a observaciones más sólidas.

Así, pues, la propuesta es la de dar un giro a este concepto y generar la idea de un Lector Modelo que se desprenda un poco del texto, pero sin abandonarlo. Porque si vamos a considerar la posibilidad de un Lector Modelo *para* ChatGPT4, la abstracción debe ser general, un modelo

¹⁰¹ Pimentel también se encarga de exponer esta cuestión en su teoría, sobre cómo el Lector Modelo se puede prever. En lugar de Lector Modelo, ella utiliza el concepto de lector implícito. “El perfil de este lector orienta al autor en todos los niveles de la escritura del texto; desde la cantidad de información descriptiva y narrativa que ha de ofrecerle (o negarle), hasta las referencias a códigos culturales que el autor supone ‘compartidos’, o bien un saber recóndito que aludir a él es ya una invitación (si no es que una orden perentoria) a salir del texto en cuestión para completarlo con otros textos. Pero en todo caso, esa imagen del lector implícito es susceptible de observación y de inferencia *a partir de las estructuras narrativas...*” (174-175; las itálicas son nuestras).



per se, una aplicación de Lector Modelo que compartirían todas las IAG de Transformador y que en todas se podría verificar. Si esto es posible es debido a que hemos venido trabajando sobre la base de modelos: tanto en la creación del cuento como en las definiciones de IAG. Hay una razón específica por la que se procedió con cautela en estas decisiones: no sólo nos permitiría dar más vigencia a los métodos de esta tesis pese a los avances de los modelos GPT, sino que nos otorga la libertad de ahondar en estas formulaciones generales con el concepto de Lector Modelo con base a lo formulado sobre las autorías empíricas, y así alcanzar conclusiones que apliquen no sólo para el cuento planteado en esta tesis, si no para muchas de las obras generadas donde las IAG tengan la mayor parte del control creativo, e incluso para las que no.

Retrotrayendo nuestras conclusiones previas, las de que el autor empírico que deviene de la unión de fuerzas de un usuario humano y de una IAG, diremos que *ChatGPT4 va a formular un Lector Modelo determinado por, antes que nada y principalmente, el nivel de la intervención del usuario*. Esto va a determinar el camino creativo a seguir, lo que tendrá un impacto enorme en formar una mejor “imagen” de ese Lector Modelo¹⁰². Ahora, ya que hemos profundizado en esta investigación en el nivel beta, continuaremos en consideración de este nivel, pero sin olvidar que estos hallazgos aspiran a ser generales, más amplios. Además, si se revisa el caso lo suficiente, concluir sobre la forma de este LM en un nivel beta no sería muy distinto a la forma que adopta en un nivel alfa o incluso gamma.

Dicho esto, el LM de ChatGPT4 *está determinado por el proceso de retroalimentación creativa*. Esto quiere decir que el proceso de *brainstorming* y de iteración tienen mucho que ver no sólo con el derrotero que tome el texto, sino con su *calidad*. Véase en el experimento llevado a

¹⁰² En este punto, esta hipótesis podría considerarse como válida asimismo para una hipótesis sobre el Autor Modelo. Y es correcto, pero sería una hipótesis incompleta. Reiteramos que este abordaje estará completo solo al revisar los apartados del capítulo 2.



cabo: se le pidió a ChatGPT4 que agregara un detective. Las marcas semánticas del semema |detective| tuvieron que haber llevado a un recorrido semiótico que al final estableció un nuevo derrotero para el cuento, hecho que se refleja fuertemente también en el Autor Modelo, ya que, según la lectura que se hizo de este, el lector debía ir de la mano con este personaje para desentrañar la clave del relato. *De no haber hecho nosotros la intervención, el Lector Modelo habría sido distinto:* en el nivel beta agregamos esta disyunción a la trama, pero la falta del detective habría llevado el ejercicio a un nivel alfa, lo que, insistimos, *altera la forma del LM*. En síntesis, *el LM de una IAG siempre va a depender estrechamente del proceso de retroalimentación creativa, de la calidad de la interacción del usuario con la IAG, un poco más que de la cantidad*. El caso del detective aumenta la intervención del usuario casi paralelamente en lo cuantitativo como en lo cualitativo, pero la disyunción que solicitamos para la trama es importante más bien cualitativamente (podría haber intervenciones cuantitativas que no afecten demasiado en lo cualitativo, y viceversa. Este caso es este último). Y ya que el LM se formula por la visión del autor empírico general, y este es una sumatoria de acciones, por eso argumentamos que el LM de ChatGPT4 depende, sí, una vez más del prompt, pero sobre todo del proceso de refinamiento del texto, de la curaduría de este, y aun verificable también por la ausencia de todo acto de curaduría (lo que degradaría la IU a un nivel alfa).

No es complicado el entender por qué el LM está ampliamente conectado con este proceso: al ser las IAG agentes racionales y que pretenden llevar a cabo su trabajo de la manera óptima, que pretenden ejecutar la instrucción del usuario hasta su resultado óptimo, significa que *el proceso de cooperación usuario-máquina se va a reflejar de muchas maneras en el texto*. El hallazgo, repetitivo a lo largo de la investigación, podrá parecer una obviedad, pero no lo es: es una conclusión que va aflorando cuando se emiten cuestionamientos sobre por qué la IAG hizo de tal



modo qué cosa. Si bien lo hace por su forma de operar y comprender su mundo, la intervención del usuario es el factor principal antes de que la IA intervenga con sus propios bagajes de conocimiento. Como señalamiento extra: si esto es realmente así, y la IU es tan importante, esto permitirá hacer rastreos más específicos de las huellas que han dejado tanto usuario como máquina, pero sólo si se conoce el proceso total de trabajo que va desde el *brainstorming* hasta las correcciones.

También debemos agregar para finalizar que, el Lector Modelo de ChatGPT *está determinado por las capacidades semánticas rigurosas del Transformador, y luego por las bases de datos de donde se alimenta*. Para el primer caso, volvamos sobre el ejemplo donde el semema detective se remite a semas muy cercanos. En términos de lo que implicaría romper el esquema semántico de este método tan rígido, es algo que no podría lograrse, y esto le impediría al Transformador lograr luego otro tipo de efectos. Como agentes racionales, y con un funcionamiento tan sistemático en la semántica, que deviene del mecanismo del *embedding*, como discutiremos más adelante a fondo, sería imposible ligar cualquier sema de |detective| a algo más alejado de él en espacio vectorial, que remita a un efecto “original” y hasta creativo. Esto es una problemática mayor en el tema de la creatividad de las máquinas, y posiblemente un hueco teórico en la materia. Luego está lo que hay en las bases de datos de una IAG como ChatGPT: estas bases son una guía de estudio, donde posiblemente se establecen fórmulas, o códigos, sobre lo que es, en este caso, el género policíaco. La IAG atiende estos códigos precisamente para *replicarlos*, no para aprenderlos y modificarlos, no para establecer invenciones, sino, reproducciones, calcos, intertextualidades, facsímiles.

Se repite, sintomática, la idea: *ChatGPT4 es una máquina obediente*. Quizá demasiado obediente para lo que a las finalidades de la creatividad literaria respecta. Elaborar textos más



creativos implicaría: a) *un fallo semántico en el proceso de generación que desembocara en un extrañamiento demasiado involuntario y difícil de ignorar*¹⁰³, b) *un acto de rebeldía por parte de la IA*, donde, en lugar de seguir al pie de la letra el contenido del prompt, como bien sabe hacer, optara por otros derroteros más interesantes según su “criterio”, algo que sabemos que no puede suceder al menos de momento, y c) *una mayor intervención del usuario*, en un nivel delta o épsilon, pero de un usuario con competencias más allá de las básicas para la narrativa. Ampliemos un poco estos posibles escenarios. En todo caso, cuando sucede a) los resultados suelen ser más interesantes —si bien no dejan de ser un error de la máquina—. Porque la literatura reposa una buena parte de su relación con el lector en ese distanciamiento, en ese extrañamiento de la realidad, y muchos lectores, incluso los más asiduos, esperan ser sorprendidos, que se rompan las reglas del género. Si bien otros prefieren, por supuesto, seguir al pie de la letra a su Autor Modelo, y ser no sólo Lectores Modelos, sino los lectores más sumisos y crédulos posibles, postura que encantaría a las obras cerradas. El inciso b) evoca otro tipo de posibilidades, más próximas, tal vez, a la ciencia ficción, y, por ende, más problemáticas. Pero de ello no nos ocuparemos aquí. El inciso c) por su parte, lo comentaremos más a fondo en el apartado 2.3. Solo hemos de formular la pregunta que nos deja esta posibilidad y que ya nos hemos hecho: si un usuario tuviera que intervenir más del 50% en el proceso de creación, ¿cuál es entonces el papel de la IAG? La pregunta es, por supuesto, retórica.

En síntesis, ChatGPT4 establece su LM tanto por la intervención del usuario como por sus capacidades operativas, las cuales, al ser parte integral de cualquier modelo de IAG de transformador, entendemos que este tipo de conclusiones tienen un alcance general a todos estos tipos de inteligencia artificial, y no focalizado únicamente a ChatGPT4.

¹⁰³ Es en estos casos cuando ChatGPT suele “inventar” sus respuestas. Los casos, por ejemplo, de bibliografía. Véase en 3.2 sobre las alucinaciones y la *bullshit*.



Más sencillo es hablar del Lector Modelo de “El Ritual de Medianoche”. De acuerdo con la estructura del cuento, podemos definir el LM, retrotrayendo también algunas reflexiones ya esbozadas en el apartado del AM.

Sencillamente, el Lector Modelo que el cuento prevé es *un lector que caiga en la trampa de creer, primero, que el asesinato fue llevado a cabo por un ente fantasmal, en lugar de por uno de los participantes del juego*. Este es al menos el punto más importante. Por eso la relevancia de que el cuento inicie con una descripción sistemática de los personajes, de los perfiles de cada uno y de sus ocupaciones: es una indicación, poco sutil, al Lector para que pueda barajar opciones. Porque como ya se dijo antes, el LM de este cuento debe convertirse en el mismo detective. Debe seguir los pasos del detective Marcos y, de hecho, adelantarse a sus conclusiones, dado que el Lector Modelo cuenta con esta información sobre los personajes mucho antes que este. El Autor Modelo deja entrever que es importante que el Lector Modelo conozca su profesión, para que luego los eventos puedan tomar sentido. El LM acepta la propuesta de su Autor Modelo. En el segundo capítulo, donde se sucede el asesinato, el cuento quiere ser leído de modo que los eventos aparentemente sobrenaturales le den al LM la indicación de que un ente paranormal está detrás del hecho. Debe, pues, como Ana, quedar convencido de ello. A continuación, seguirá los pasos del detective, formulará sus hipótesis sobre “*¿quién o qué mató a María y por qué?*” e irá haciendo sus tiradas interpretativas.

Para el final del cuento, el LM ha resuelto el caso, el asesinato, y, según lo propone su AM, deberá comprender la reflexión final, deberá aceptarla. La reflexión moral hacia el final hace que



el LM de “El Ritual de Medianoche” resignifique todo el cuento, y otorgue a este un significado más elevado, uno moral. El LM debe entender que, con esta reflexión, el cuento, si bien es simple, va más allá de la historia en sus páginas: le ha dado una reflexión para la vida diaria. Al menos así desea su AM que lo haga.

Otras generalidades son que el LM debe ser un hablante del idioma español y debe estar familiarizado con los códigos básicos de la poética Aristotélica (introducción, desarrollo, clímax, anagnórisis, conclusión) y del género policial, que formulados a manera de macroproposición se condensan en: “un misterio que debe ser resuelto”, y que como cuadro intertextual abre un abanico de personajes y situaciones, de las que se destacan el detective entre los personajes, y el asesinato entre las situaciones.

Por tanto, el LM de “El Ritual de Medianoche” es un lector *que cae en la trampa de Sergio, pero que se convierte en el detective Marcos, siguiendo huellas y pistas que revelan la verdad, y que valora la fuerza de esta ante la mentira y el engaño*, pues “la verdad siempre encuentra su camino”.

Sin embargo, podemos ir un poco más allá y formular algo sobre un hecho meramente accidental, o que se da por defecto, que depende más de la IA (en un nivel beta) que del usuario. Si bien esta definición de Lector Modelo por sí sola suena lo suficientemente bien, la realidad es que el LM de este cuento debería ser, realmente, *un lector demasiado ingenuo*. El cómo específicamente es una formulación que podremos continuar en el siguiente capítulo, ya que nos conducirá a concluir sobre la calidad de la cooperación interpretativa de ChatGPT4 con el lector.



Capítulo II. El proceso de cooperación interpretativa en el texto de ChatGPT4

—¡A mí me lo vas a decir!

Que quede entre nosotros,

¡no sabes lo difícil que es ser

un modelo de Inteligencia Artificial!

Los límites de la interpretación, Umberto Eco

2.1 La cooperación interpretativa con el lector en “El Ritual de Medianoche”: obra abierta, obra cerrada

“*¿Qué o quién mató a María y por qué?*”, es el eje central de la cooperación interpretativa en el cuento generado con la IAG. Hasta ahora hemos recolectado las definiciones de AM/empírico y LM/empírico establecidas por ChatGPT4 en su texto. Las conclusiones recabadas nos han llevado a entender que el Lector Modelo de ChatGPT depende principalmente de la intervención del usuario. Este hallazgo, junto con el de la conclusión sobre la generación del lector ingenuo, explorados a través del eje central, nos permitirá ahora formular la calidad de la cooperación interpretativa en “El Ritual de Medianoche”: si es una *obra abierta*, que invita a la interpretación y a la reinterpretación, que atrapa al lector y lo guía a hacer sus propias conclusiones, si le proporciona varias vías de acceso para jugar un rato con esa interpretación; o si es una obra hermética y que empobrece la relación con su lector, llegando a incluso imponer su lectura, su contenido, empobreciendo sus posibilidades comunicativas. Lo mismo que antes, extraer estas conclusiones nos permitirá entender la creación literaria de las IAG en un modo general.



Ahora bien, antes hemos expresado ya que no todo texto permite esta apertura interpretativa, pero todo texto abierto postula, de cierta forma, su apertura a la interpretación (si bien también puede postular su hermetismo): “... *un texto es un producto cuya suerte interpretativa debe formar parte de su propio mecanismo generativo*: generar un texto significa aplicar una estrategia que incluye las previsiones de los movimientos del otro...” (Eco, *Lector in fabula* 79; itálicas son del autor). Y esto es totalmente válido para el texto de ChatGPT4. Se lo puede ver en ya desde la forma en que dispone inicialmente de sus personajes. Pero la labor de la que nos encargaremos aquí es más bien la de definir cómo este cuento desenvuelve esta apertura interpretativa, si lo hace de modo que se demuestra como texto cerrado o abierto en la totalidad del proceso de cooperación interpretativa. Aun cuando todo texto tenga implícitas las pautas para su interpretación, no significa que hayan sido dispuestas por su autor, como sí que el lector podría encontrarlas. Además, como menciona Eco en el *Lector in fabula*, una distinción importante entre obra abierta y cerrada tiene que ver también con “la intensidad y la vivacidad de la cooperación” (171), cuestión que, como también señala, puede representar un criterio para la valoración estética de la obra (171).

Para dar con estas estrategias debemos puntualizar más. Y esto se logrará volviendo sobre los conceptos antes explorados. Hemos hablado por separado de AM y LM. Tan solo en momentos puntuales se han visto encontrados o atravesados el uno por el otro. Pero según establecimos en el marco teórico, “...la noción de interpretación supone siempre una dialéctica entre la estrategia del autor [AM] y la respuesta del Lector Modelo” (86). Del mismo modo, más adelante dice: “... la cooperación textual es un fenómeno que se realiza entre dos estrategias discursivas, no entre dos sujetos individuales” (91), es decir, no se realiza entre un autor empírico y un lector empírico. De modo que ahora debemos entender la relación entre AM y LM.



Una obra abierta admitiría múltiples lecturas, o múltiples formas de llegar al tema de la obra, como un ejercicio estimulante para sus lectores, o cuando menos no tratará de forzar su entendimiento. El autor de una obra que opera como abierta decide “hasta qué punto debe vigilar la cooperación del lector, así como dónde debe de suscitarla, dónde hay que dirigirla y dónde hay que dejar que se convierta en una aventura interpretativa libre” (84). Estas cualidades apelan, finalmente, a una libertad interpretativa dentro de la obra, libertad atribuible a esta poética literaria contemporánea de la obra abierta, de no hacer un uso obsesivo de una obra, de no guiar al lector a donde el autor deseé sino de confiar en el acto pragmático que es reconocer la autonomía de ese lector y las diferencias con su enciclopedia, con sus códigos. De este modo, el lector podrá rescatar un mayor provecho de la lectura, de la interpretación de la lectura en cuanto que remite a lo que se decía al inicio de esta tesis: que el acto de lectura es una configuración de la experiencia, una metáfora epistemológica. Esta experiencia puede ser una autónoma sólo si el texto así lo permite, o una instruccional, en su defecto. Dicho esto, toca definir concretamente si lo que produjo ChatGPT4 es un texto cerrado, que prevé un LM inequívoco y obediente, o si es un texto abierto. Parte de la respuesta ya la tenemos.

De entrada, una pregunta: ¿qué tantos espacios en blanco se dejan para que el lector actualice el texto de “El Ritual de Medianoche”? ¿Qué tanta autonomía se le otorga al lector en este sentido? Volvamos ahora sobre el eje central de la cooperación. “*¿Qué o quién mató a María y por qué?*” Cuando el cuento establece un título, “El Ritual de Medianoche”, y un subtítulo en su primer capítulo, “El juego Comienza”, el lector, como señalamos al comienzo del apartado el Autor Modelo, empezará a formular un *topic*, o tema, pero este no tomará una forma concreta hasta que se sucedan los primeros dos capítulos. El tema es actualizable conforme a la lectura lineal del texto. Pasados estos capítulos, el ingenuo lector de ChatGPT4 llegará sin mayores dificultades a



la formulación del tema del texto: asesinato misterioso. Esta es una de las primeras bases de la cooperación interpretativa: el “descifrar” de *qué* se está hablando. El Lector Modelo ha entrado, entonces, en acción: debe resolver, con la lectura, este asesinato misterioso. El Autor Modelo se encarga de comunicar varias señales a su LM sobre este asesinato misterioso, porque ya que todo texto quiere ser leído para cobrar vida, tiene que otorgar ciertas pautas: los terrores de Ana, los silencios abrumadores, los hechos paranormales. Con esto, el AM está pidiendo de su LM que conjecture, como ya se dijo, que algo paranormal está acechando la casa rural. Sigue entonces la muerte de María, sobre la cual no se ahonda en descripciones de momento: solo en el capítulo tres se mencionan marcas en el cuello, pero no se sabe de qué, una imprecisión de información que más que atraer la atención del lector podría confundirlo. Así, pues, la pregunta de “*qué mató a María*” está activa en el Lector Modelo.

Este primer trabajo interpretativo no es sino básico. Una serie de lecturas de expresiones que remiten a sus contenidos como en todo sistema semiótico. Una cadena *fuerte* de sememas guía al lector a la creencia en el caso paranormal: |Ritual|, |Medianoche|, |Nerviosa|, |Ouija|, |circunstancias misteriosas|, etc. Y entonces el lector, típicamente, debería enfocarse únicamente en mantener activa esa pregunta: *¿qué mató a María?*, y anestesiar otras posibilidades. Porque, como dijimos, esto es lo que el Autor Modelo desea de su Lector Modelo, y este, como es ingenuo y obediente, acepta seguir esta trama. Sin embargo, aquí ya tenemos un primer desencuentro con las expectativas entre AM y LM: *ningún lector es tan ingenuo como para ignorar las marcas tan evidentes del personaje de Sergio*. Ya desde el capítulo 1, el AM establece pobemente una rotunda brecha entre este personaje y los otros, si bien no basta para sospechar que él sea el culpable: aquí todavía puede esperarse que el AM esté tratando de llevar la atención a un lugar equivocado, a Sergio como el asesino, para evitarle ver al culpable real. Pero justo después del asesinato de María



se vuelve a hacer esta mención de Sergio como inescrutable y sospechoso: “Sergio permanecía en el fondo, observando en silencio” (capítulo 2). Lo ideal para abrir la obra sería entonces que se encontraran en el camino del detective huellas que reforzaran ambas teorías: que un ente paranormal mató a María/que Sergio mató a María. Pero no es así: el único respaldo que se tiene es que Ana vuelve a hacer énfasis en sus miedos, hecho que no es suficiente, sobre todo cuando se vuelve a hacer la mención de Sergio que “como siempre, era el más frío” (capítulo 3)¹⁰⁴. Hasta este punto, el AM, satisfecho y seguro de sí mismo, continúa guiando a su LM, estableciendo, quizá, una falsa apertura en la obra: pretende darle espacio al LM para que dude, pero, ya que aceptamos que este no puede ser tan ingenuo, aventurará al finalizar el tercer capítulo que, en efecto, Sergio es el asesino, no hay misterio. Volviendo, entonces, a la pregunta inicial: ¿qué tantos espacios libres ha dejado el texto para navegarlos? Al parecer demasiado pocos, o sólo uno. La tensión muere antes de la mitad del cuento. Esta tirada interpretativa rompe entonces la dialéctica de AM y LM y revela la cerrazón de la obra: su fallo en emitir varias disyunciones posibles para el *topic* de la historia, para la tirada interpretativa principal¹⁰⁵. Y así, lo que Eco sugiere para las disyunciones de una trama, que, “Naturalmente, el autor debe escoger después la solución menos probable, porque lo que busca el poeta es producir asombro” (169), queda fuera de las

¹⁰⁴ En el *Lector in fabula* se habla de un término cuya ausencia en este punto en específico funge como criterio para la valoración estética de la narrativa: “un texto narrativo introduce diversos tipos de señales textuales destinadas a subrayar que la disyunción que está por aparecer es pertinente. Esas señales se denominan *señales de suspense*: pueden consistir, por ejemplo, en una dilación de la respuesta a la pregunta implícita del lector” (159). Recordemos que la pregunta implícita principal es “¿qué o quién mató a María y por qué?”. Esta pregunta no tiene realmente actualizaciones pertinentes tras de que es planteada.

¹⁰⁵ En su bibliografía, Eco se apoya constantemente del concepto de “paseo inferencial”, concepto que congrega las previsiones, inferencias y mundos posibles que el lector va creando durante el proceso de cooperación con el texto. Un paseo inferencial implica salir del texto y ahondar en un modo de conocimiento o cuadro común que le permita reforzar esas inferencias. Muchas veces esa información viene dada por la experiencia, una que puede ser de índole intertextual, por ejemplo: “si la fabula le dice ‘x realiza tal acción’, el lector aventurará: ‘como cada vez que un x realiza tal acción suele producirse el resultado y’, y concluirá: ‘entonces, la acción de x tendrá como resultado y’” (166). Esta es otra de las razones por las que al lector empírico no le cuesta llegar a la conclusión sobre quién es el asesino mucho antes de lo que desearía su AM. El AM pudo haber “aprovechado” esta inferencia común para jugar con el lector, sin embargo, tal uso de estrategia estaba lejos, al parecer, de sucederse.



posibilidades del texto. Tras de la revelación del asesino, podría forzarse un poco el texto para hacerlo hablar, jugar un poco con él, pero ello no alteraría nuestros resultados, ya que, hay que recordar, lo que cambia entre obras abiertas y cerradas es “la intensidad y la vivacidad de la cooperación” (171): la antes mencionada dialéctica entre AM y LM, y esta no destaca ni por su intensidad ni vivacidad. Ir más allá implicaría simplemente ejercer violencia al texto para abrirlo y avenir su uso para muchas otras cuestiones, como se verá más adelante.

Pero supongamos que por alguna razón el LM pueda seguir dudando. Podemos concederle al texto todavía ese punto a favor. Tal ejercicio nos llevaría a la búsqueda de la calidad de la cooperación interpretativa en otros aspectos de la obra, como hemos mencionado.

Es posible preguntarse: ¿acaso el texto ayuda a construir la competencia con que quiere ser leído, como propone Eco que hacen los textos? Una primera respuesta es sí, porque es un artificio común a todo texto: un autor empírico no puede prever a su lector empírico, pero sí contribuir a crear su Lector Modelo. El cuento del Ritual, o el AM, instituye sus competencias ya desde su léxico, desde esa ausencia de descripciones en privilegio de secuencias de acontecimientos que construyen apresuradas por llegar al conflicto. La forma en que lo hace, entonces, habla mucho de *qué tanto dialoga con el Lector Modelo y qué espera de este diálogo*: si espera confundirlo, guiarlo a puntos que ensanchen su duda, etc. Si seguimos esta línea argumentativa, encontraremos que la mayor demostración de cómo el Autor Modelo, establece sus competencias es a través de la figura del detective. Todo el discurso que rodea a esta figura es una indicación más o menos directa para el LM sobre cómo debería leer el texto, y por tanto debería aspirar a generar las competencias necesarias. Digamos, en este caso en particular, que hay una posibilidad de compartir con el Lector una enciclopedia¹⁰⁶ de términos relativos a la investigación, a la criminalística, al arte de la

¹⁰⁶ “Otro tipo de señales al narratario lo constituyen las referencias a objetos o entidades en el mundo del extratexto cuyo ‘conocimiento’ y ‘re-conocimiento’ por parte del lector presupone el narrador: ‘uno de esos objetos que se



deducción (que es semiótico, de hecho), y sin embargo, no lo hace. Una enciclopedia que demuestre las audacias del investigador para desmontar cualquier fenómeno paranormal y hacer ver que las pruebas remiten a un autor humano brilla por su ausencia. Y lo que es más, la enciclopedia que propone el AM del cuento es bastante endeble. Se refleja en oraciones como “Marcos comenzó a atar cabos”; sememas como “entrevistó”, “descubrió”, etc. Ahora bien, no es que exijamos de las posibilidades de los autores empíricos una riqueza léxica que nos eduje en el arte de la deducción (o la abducción, según Peirce): simplemente estamos argumentando que el texto no propone nada ni siquiera a este nivel, lo cual, finalmente, es otro criterio estético. La segunda contribución enclopédica mayor gira alrededor del léxico esotérico-demonológico. El uso del término francés del indicador de la ouija es un ejemplo. La presencia de estos términos podría figurar como un respaldo del AM para contribuir a que el LM se apegue a una primera interpretación falsa sobre lo que mató a María, sobre todo porque figuran en la primera mitad del texto. Salvo por este ligero acierto del LM, que como parte de la investigación sabemos que es una contribución directa de ChatGPT4, el resto del cuento es muy lineal en cuanto a su cooperación enclopédica. Esta ausencia de creación de competencias enclopédicas es otra forma de configurar una obra cerrada.

Para ahondar más en las cualidades de la cooperación interpretativa, podemos movernos a otra parte de la estructura del texto: el narrador. Según quedó establecido en el marco teórico de la mano de Luz Aurora Pimentel, el narrador, en sus muchas formas, es un mediador entre el mundo narrado y el lector. Y en efecto: el AM de “El Ritual...” proporciona información de más al Lector a través de su narrador. Y es aquí donde cuestionamos ahora sí el por qué esta discriminación en

compran en x parte’, ‘una de esas calles en las que...’ Como bien podrá observarse, todas estas descripciones envían señales al narratario, más que para su comprensión, para su reconocimiento, para que el lector les dé su acuerdo, y comparta su ‘enciclopedia’ con la del autor” (Pimentel 178; itálicas de la autora).



los tipos de narradores. Si el usuario no ha hecho la selección del tipo de narrador, ¿por qué le ha parecido a ChatGPT4 este tipo de narrador el ideal? La pregunta es retórica, pero en términos de lo que se refiere a la cooperación interpretativa, podemos suponer que una voz omnisciente, que de hecho no hace ningún esfuerzo por ahondar en las mentes de los personajes, sino rescatar únicamente sus emociones visibles (“Sergio, el abogado, se mantenía en silencio. Sentado en la esquina”), habría tenido como objetivo separar al Lector Modelo lo más posible de la develación del misterio. Si se describe a Sergio, el culpable del crimen, como inescrutable, se implica que ni el mismo narrador tiene acceso a la culpabilidad del actor. El LM tampoco debería, por tanto, sospechar nada de momento. Pero *la selección de esta descripción del estado del personaje ya por sí sola determina una marca al lector*: este debe de sospechar de Sergio. Lo vimos antes por la contigüidad de esta información con el acontecimiento de la muerte de María, casi como un acto correferencial. Pero el lector que propone el cuento es ingenuo, y las competencias de cualquier lector, comunes o intertextuales, podrían determinar con facilidad que Sergio es el culpable solo con la lectura de este pasaje, solo con plantearse la pregunta de “*¿por qué el narrador quiere que sepa esto de este personaje?*” (Para el lector es el narrador, pero como teóricos de la literatura sabemos que no es otro artificio que el Autor Modelo estableciendo selecciones de información narrativa y descriptiva). De modo que la selección de narrador que se propuso ChatGPT4 para aventurarse en el género policiaco *pretende* apartar al Lector Modelo de realizar las inferencias adecuadas. No obstante, ningún lector empírico ha de mostrarse necesariamente tan ingenuo como para no sospechar más allá de lo discursivo. Y si el Lector llega a formularse esa pregunta, de por qué el narrador está proveyéndole de esa información, es porque está entrando en un acuerdo natural de interpretación. Sólo recuérdese que esto es natural a todo texto, a todo signo, y no precisamente un acierto de la obra.



El fallo en este seguimiento del narrador que el AM propone se da por lo que menciona también Pimentel: que el lector “tiene una posición privilegiada frente al mundo narrado, ya que es él quien hace convergir todas las demás perspectivas [de los personajes y el narrador], acumulando de este modo formas de significación más complejas” (173). Al no haber ni siquiera un juego de perspectivas, sino reacciones planas y delatoras, un texto tan cerrado y transparente como lo es el cuento de la IAG no hace sino evidenciar esta posición privilegiada del lector: el texto no representa ningún desafío para él.

Otra consideración que hace de este texto una obra cerrada es que no es sólo que la cooperación con el LM sea débil —el solo hecho de accidental o indirectamente proponer un Lector ingenuo es también una forma de cerrar las posibilidades de la obra—, poco enriquecedora y que el texto parezca más bien querer llegar con urgencia a su final, sino que importa mucho lo que sucede en este mismo. Ya hemos mencionado antes cuál fue el planteamiento final del AM, y ahora hemos de volver a esto. Sabemos por la teoría de Eco que un texto es cerrado también cuando “... en casos de extrema pedantería, de extrema preocupación didáctica o de extrema represión el texto se complica con redundancias y especificaciones ulteriores” (76). Y aunque en el caso de este cuento las preocupaciones “didácticas” no parezcan tener demasiada relevancia, lo sorprendente es que existe esta enfática función didáctica, que en ningún momento se sugirió ni de soslayo. Quizá esto es lo más interesante en cuanto a explorar las particularidades creativas de las IAG. Sin embargo, para mantenernos en el texto, podemos únicamente concluir que la obra es cerrada continuando la cita de Eco del mismo párrafo: “a medida que pasa de la función didáctica a la estética, un texto quiere dejar al lector la iniciativa interpretativa, aunque normalmente desea ser interpretado con un margen suficiente de univocidad. Un texto quiere que alguien lo ayude a funcionar” (76). En este caso, como es evidente, lo que se da es lo opuesto: se pasa de la función



estética a un cierre didáctico, moralizante, que sin duda debe sorprender a cualquier lector empírico avisado y cuando menos un poco informado en lo que respecta a las inteligencias artificiales generativas.

De modo que no podemos sino volver a hacer énfasis en el fallo: a manera de juicio tanto crítico como estético, *ChatGPT4 y su coautor empírico, en este nivel de creación beta, produjeron una obra cerrada, que trunca bastante la cooperación interpretativa con el lector*. A riesgo de parecer demasiado insistentes, remitimos a la metodología que hemos utilizado para la creación de este cuento: esta nos permite observar que el carácter de esta conclusión es general, va más allá de “El Ritual de Medianoche”. Este cuento es una producción modelo para dilucidar cómo serían otros posibles cuentos de un ChatGTP4 utilizado para bajos niveles de intervención del usuario, y estas observaciones deberían de poder trasladarse a trabajos incluso más amplios, algunos de los cuales se exhiben en la comunidad de videos y creadores de contenido, por ejemplo, que hemos rescatado para la metodología.

“Nada más abierto que un texto cerrado” (83), dice Eco sobre el carácter de estas obras. “...aparentan requerir nuestra cooperación, pero subrepticiamente siguen atendiendo sus propios asuntos; son textos ‘cerrados’ y represivos” (304). Y esto no implica sino que un lector común se vuelva propenso a hacer el uso que quiera de estos cuentos. Nada le impediría, si bien las posibilidades no son más que infructíferas, lúdicas, argumentar que de algún modo María sí fue asesinada por un fantasma, o que todos estaban muertos, o que Marcos fue el fantasma que asesinó a María, y un largo y lúdico etcétera. Pero esto es ya, justamente, jugar con el texto, abrirlo no sin violencia: nada más abierto que el texto cerrado. Porque en un texto cerrado,

... una, y sólo una, será la hipótesis correcta. A medida que la fabula [historia] se va realizando y se va disponiendo a lo largo de su eje temporal, pone a prueba las



anticipaciones, excluye las que no corresponden al estado de cosas al que desea referirse y, por último, traza una especie de línea cosmológica continua según la cual lo que ha acontecido ha acontecido y lo que no ha acontecido ya carece de importancia ... este tipo de fabula es cerrada, por cuanto no permite (al final) alternativa alguna y elimina el vértigo de las posibilidades. El mundo (de la fabula) es el que es (170).

Y esto es lo que se propone en “El Ritual de Medianoche”. Cualquier otra posibilidad es ejercer un tipo de violencia en el texto.

Y cuando este uso de los textos se da, aparece en la cultura un fenómeno que pretende volverlos, junto con las IAG, algo que no son. El fenómeno del *AI Washing* de lleno en la literatura. Es lo que se ha intentado hacer en esta comunidad que mencionamos antes, donde impera una necesidad aparente de acelerar la producción creativa, con visibles detrimientos en la calidad de los textos¹⁰⁷. Más fructífero que solo reflexionar sobre esto resultaría preguntarnos qué papel estamos otorgando, o se nos pide otorgar, a estas tecnologías en nuestras vidas, en las áreas en las que ahora sabemos que incursiona.

Hemos establecido que “El Ritual de Medianoche” es una obra cerrada. Del mismo modo que hicimos con los conceptos de Autor Modelo y Lector Modelo, diremos que las inteligencias artificiales generativas *tienen una tendencia a producir obras cerradas*, a establecer un vínculo endeble y reticente con su lector durante la cooperación interpretativa.

¹⁰⁷ Véanse los videos de YouTube “I Used ChatGPT to Write a Book. And I Published it | ChatGPT for Authors | Writing a Novel With AI”, <https://www.youtube.com/watch?v=R59rYTEe9Sw>; y “How to Write a Book in a Day with AI (My Approach)”, https://www.youtube.com/watch?v=OTa6a7NGP_Y.



Pero más allá de solo observar este resultado a partir del cuento generado con la IAG, es necesario reforzar este planteamiento con evidencia que está implícita en el funcionamiento propio de las IA. Porque es desde este punto específicamente que proviene esta tendencia a las obras cerradas. Es a través de esta dilucidación que terminaremos de dar forma a la estrategia textual que es el Autor Modelo, como dijimos antes. Y es que la parte del funcionamiento de las IAG que sugerimos que fomenta la cerrazón de sus obras literarias, y que es la causa de una calidad plana en sus obras, en su lenguaje, en sus tramas, en su contenido general, está más allá del mero acatamiento de órdenes y su ejecución óptima. Lo que argumentamos es que su mecanismo semántico-vectorial juega una parte importante en la calidad de estos contenidos y, por ende, de la creación de obras con poca comunicación con su lector: cerradas.

2.2 La obra cerrada y el embedding vectorial

2.2.1 Fundamentos del embedding vectorial

La inclinación a la obra cerrada —así como a muchas otras cosas— está directamente fundada en el funcionamiento de las IAG de Transformador. En el estado del arte crítico obtuvimos como conclusión que algunos de los fallos mayores en los textos creativos de las inteligencias artificiales se debían a la ausencia de una comprensión semántica real de los textos, y, por ende, a la ausencia, todavía más notable si se quiere, de una visión pragmática de los textos. Esta conclusión permeaba a inteligencias artificiales generadoras de texto, pero no se verificó en ningún momento directamente con los LLM. Las conclusiones, sin embargo, no distan demasiado¹⁰⁸. El problema de la creatividad de las máquinas, de la baja calidad en la creatividad y sus derivados,

¹⁰⁸ Es posible que inteligencias artificiales previas hayan operado bajo un modelo de algoritmo de tipo regresión lineal, encargado de predecir valores entre variables, modelos que operan mucho mejor en campos de análisis de datos, como sectores financieros, por ejemplo. El concepto sigue siendo el mismo: la transformación de datos a números, un proceder propio de la lógica matemática.



puede puntualizarse en la nula comprensión semántica y pragmática de los textos, pero en lo que respecta a las IAG de Transformador, el problema puede puntualizarse todavía más cuando se explora el mecanismo de los *embeddings*.

Un *embedding* o incrustación vectorial se define de la siguiente manera, en el contexto de su aplicación dentro del aprendizaje de máquinas, o *Machine Learning* (ML), una subrama de la inteligencia artificial: “Los *embeddings* vectoriales son representaciones numéricas de puntos de datos que expresan diferentes tipos de datos, incluidos datos no matemáticos como palabras o imágenes, como una matriz de números que los modelos de aprendizaje de máquinas (ML) pueden procesar¹⁰⁹” (Bergman, Dave, Stryker, Cole, “*What is vector embedding?*”). Es posible todavía puntualizar más la definición. La siguiente define los *embeddings* en el contexto de la lectura del contenido de un texto y amplía la información sobre su uso:

... representaciones del significado de las palabras, llamadas *embeddings*, directamente a partir de sus distribuciones en los textos. Estas representaciones se utilizan en todas las aplicaciones de procesamiento del lenguaje natural que utilizan el significado ... Encontrar estas formas auto supervisadas de aprender representaciones de la entrada, en lugar de crear representaciones manualmente mediante ingeniería de características, es un objetivo importante de la investigación en PLN.¹¹⁰ (Jurafsky, Martin, *Speech and Language Processing: An Introduction*... 101).

¹⁰⁹ Original: “*Vector embeddings are numerical representations of data points that express different types of data, including nonmathematical data such as words or images, as an array of numbers that machine learning (ML) models can process*”.

¹¹⁰ “... *representations of the meaning of words, called embeddings, directly from their distributions in texts. These representations are used in every natural language processing application that makes use of meaning ... Finding such self-supervised ways to learn representations of the input, instead of creating representations by hand via feature engineering, is an important focus of NLP research*”.



Esta forma de procesar la información encuentra su base conceptual en la lógica matemática, disciplina fundamental de la ciencia de la inteligencia artificial¹¹¹. Esto implica que “Cualquier dato con el que opere un modelo de IA, incluyendo datos no estructurados como texto, audio o imágenes, debe expresarse numéricamente¹¹²” (Bergman, Dave, Stryker, Cole, “*What is vector embedding?*”). Pero esta forma de estructurar datos, de estructurar numéricamente la realidad que describen estos datos, no viene dada de manera aleatoria. Los valores numéricos son asignados de primera instancia de manera aleatoria en una matriz vectorial, pero en el proceso de aprendizaje de la máquina estos se van ajustando con base a información nueva (3Blue1Brown, Transformers (how LLMs work) explained visually | DL5” 13:18), y entran entonces otro tipo de mecanismos para descifrar mejor estos datos, mecanismos que ya hemos mencionado en el apartado 1.1 (*weights* y *Attention*) y que se consideran como otorgadores de contexto, es decir, que definen las relaciones entre las palabras, para darle un mejor entendimiento a la máquina de su uso. Como sucede con la tecnología de las redes neuronales, estos procedimientos suenan sencillos en su faceta conceptual, pero en conjunto se estaría hablando de una ejecución de millones de datos numéricos, que finalmente se resuelven en el prodigo de la lógica matemática y la PLN que es el *embedding* vectorial.

Lo que sigue a este proceso es relacionar esa serie de vectores entre sí, y empezar a generar un segundo nivel de significado a ese primer nivel numérico. Esta segunda capa de significado es la que se pretende como una nueva capa semántica: empiezan a verificarse similitudes entre datos, entre conceptos, respecto a sus posiciones en el campo vectorial, y esto lleva a deducir un posible

¹¹¹ *Embedding* y *vector*, son términos diferentes dentro del *Machine Learning*. El *embedding* es, como lo dice la definición, una traducción numérica de datos, mientras que el *vector* implica la representación de esta en un espacio n-dimensional. Un *embedding* no siempre tiene que convertirse en un vector, por ejemplo. Sin embargo, incluso en el mismo campo del ML estos términos suelen ser intercambiables. Para fines de esta investigación, adoptaremos este enfoque donde los términos son considerados como sinónimos.

¹¹² Original: *Any data that an AI model operates on, including unstructured data such as text, audio or images, must be expressed numerically*.



contexto y uso de dichos datos, puesto que también se pueden definir por la cercanía en el campo de otros conceptos, por su relación con estos.

Intuitivamente, cuanto más similares sean dos puntos de datos del mundo real, más similares deberían ser sus respectivas incrustaciones vectoriales. Las características o cualidades compartidas por dos puntos de datos deberían reflejarse en ambas incrustaciones vectoriales. Los puntos de datos diferentes deberían tener incrustaciones vectoriales diferentes.

Con estas suposiciones lógicas, las incrustaciones vectoriales se pueden utilizar como entradas para modelos que realizan tareas útiles en el mundo real a través de operaciones matemáticas que comparan, transforman, combinan, ordenan o manipulan de otro modo esas representaciones numéricas

Expresar puntos de datos como vectores también permite la interoperabilidad de diferentes tipos de datos, actuando como una especie de lengua franca entre diferentes formatos de datos al representarlos en el mismo espacio [vectorial] de *embedding*.¹¹³ (Bergman, Stryker, “*What is vector embedding?*”).

Existen varios ejemplos visuales de cómo luce este proceso de *embedding*. Uno particularmente educativo ilustra las posiciones en el campo semántico de dos figuras dictatoriales. Para esto, el modelo de inteligencia artificial primero tendría que haber sido alimentado con información desde una base de datos. A esos datos se les habría agregado un valor numérico que

¹¹³ Original: *Intuitively, the more similar two real-world data points, the more similar their respective vector embeddings should be. Features or qualities shared by two data points should be reflected in both of their vector embeddings. Dissimilar data points should have dissimilar vector embeddings.*

Armed with such logical assumptions, vector embeddings can be used as inputs to models that perform useful real-world tasks through mathematical operations that compare, transform, combine, sort or otherwise manipulate those numerical representations.

Expressing data points as vectors also enables the interoperability of different types of data, acting as a lingua franca of sorts between different data formats by representing them in the same embedding space.

tendría que haber pasado por numerosos ajustes hasta que los resultados se mostraran óptimos. Suponiendo que el modelo de IA hubiera sido administrado con información concerniente a las guerras de la historia de la humanidad, particularmente las del siglo XX, tendría conocimiento sobre los países y las principales figuras involucradas en los principales conflictos bélicos y sociales. Esta información el *embedding* la habría dispuesto en el campo vectorial, para poder procesarla, para otorgarle significado a esos *tokens* recabados. Entonces, si se insertara una fórmula para este campo, que tuviera la forma siguiente, donde *E* significa *embedding*: $E(\text{Hitler}) + E(\text{Italy}) - E(\text{Germany}) = X$, esta variable tendría como resultado *aproximado* el valor de *Mussolini*.

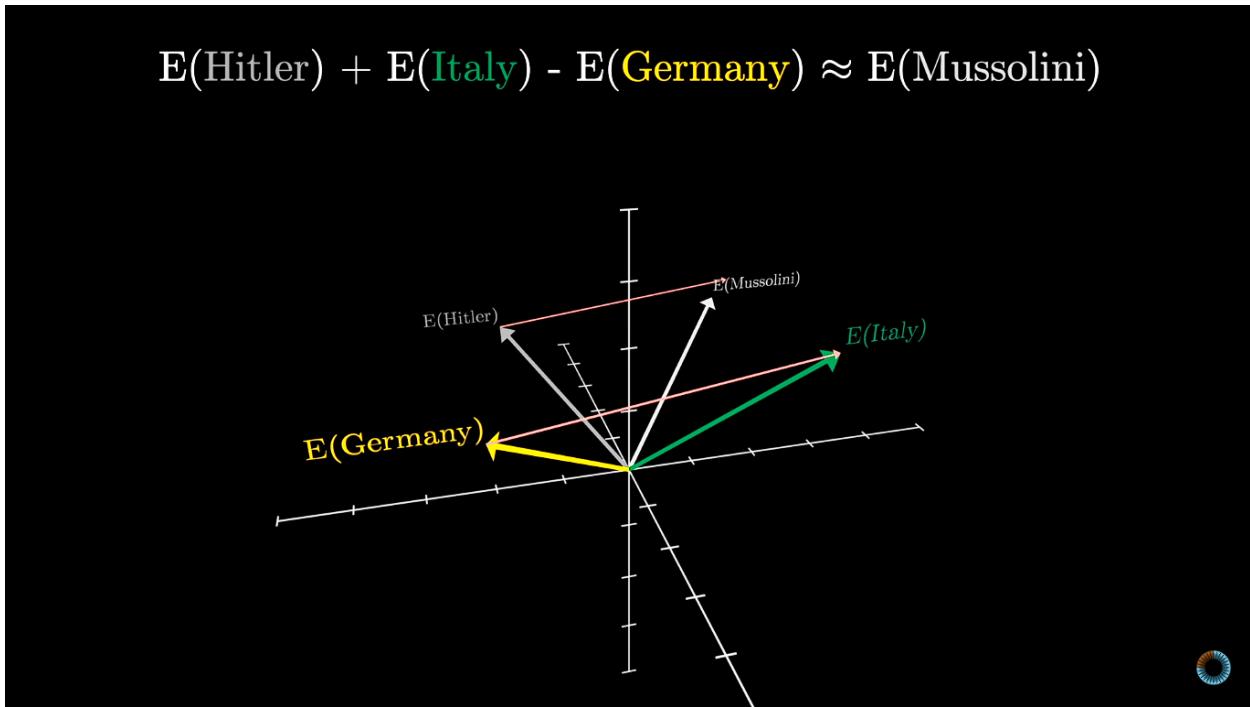


Figura 4. Fuente: 3Blue1Brown: “*Transformers (how LLMs work) explained visually | DLS*”

La figura 4 es una representación visual de la fórmula que incluye su resultado. Algo más que habría que agregar a esta conceptualización es que la dirección de los vectores (su orientación en los ejes) es importante, y ayuda a comprender la forma en que estas palabras tienen un soporte contextual. Esta forma de operar encuentra su modelo en los vectores según su aplicación en la física, cuya naturaleza se compone de una dirección, un sentido y un valor; sin embargo, no son

precisamente los mismos tipos de vectores (Bergman, Stryker, “*What is vector embedding?*”). El par de *embeddings* vectoriales E(Mussolini) y E(Italia) se encuentran orientados en una dirección similar, dado que el modelo, como resultado de su aprendizaje, ha añadido a esa orientación una capa de significado extra que podría denominarse como un “eje de orientación relacionado con lo italiano”, mientras que la orientación similar de los E(Hitler) y E(Mussolini) tendría una connotación de “eje de dictadores de la Segunda Guerra Mundial”¹¹⁴.

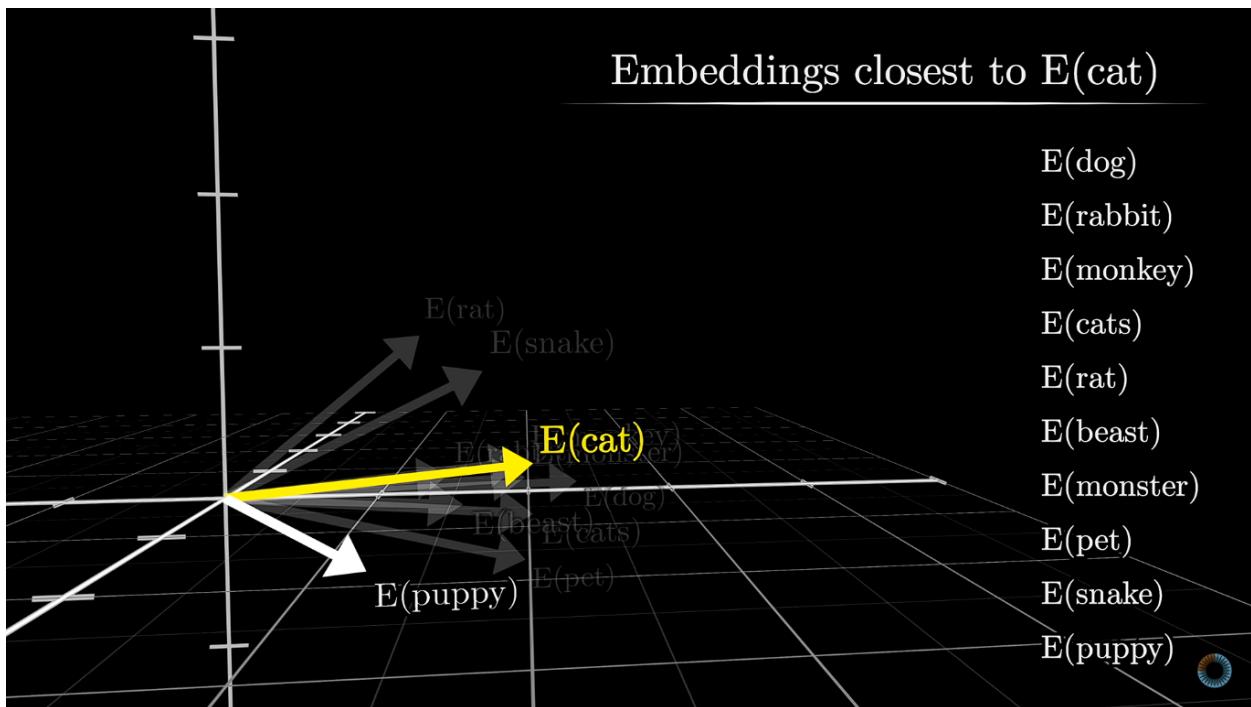


Figura 5. Fuente: 3Blue1Brown: “*Transformers (how LLMs work) explained visually | DL5*”

¹¹⁴ La verdadera complejidad del *embedding* vectorial consiste en la posibilidad de agregar miles de dimensiones para procesar la orientación (lo que hemos denominado como connotaciones o como orientaciones tal cual) de los vectores en el espacio, y establecer así relaciones entre los *embeddings*. De modo que el ejemplo que utilizamos aquí vuelve a ser meramente conceptual, pero es perfectamente útil y válido para las argumentaciones expuestas a continuación. Lo más importante es que pese a esta cantidad elevada de dimensiones, es posible establecer relaciones entre los *embeddings* vectoriales y sus orientaciones. Si esto no fuera posible, el mecanismo del *embedding* vectorial sería inservible. Cualquiera que sea el número de dimensiones en el espacio planteado, existen formas de medir las distancias entre los vectores con consideraciones y operaciones algebraicas básicas: fórmulas para la distancia euclíadiana, fórmulas trigonométricas para los cosenos y el producto punto vectorial (Bergman, Stryker, “*What is vector embedding?*”).



En la figura 5 se pueden visualizar no solo este factor de las orientaciones de los vectores, que en el caso mostrado la orientación bien podría denominarse bajo una connotación de “animal”, o una conjunción entre “mamíferos” y “mascotas”, sino también diferentes tipos de datos similares a aquél de E(Gato) agrupados en cercanía. Únicamente parecen descuadrar los vectores de rata y serpiente, pero lo importante es que pueden encajar en una orientación semántica de “mascota” y un cuadro contextual que los aproxime a “gato”, de manera que no se rompa la lógica de su cercanía con el *embedding* gato. Incluso el *embedding* rata podría tener una relación cercana con gato y con serpiente por una connotación de depredación. A este proceso le sigue luego el del *unembedding*, algo así como la “desencriptación” de los datos y que opera bajo la operación de una distribución de probabilidad, que define qué palabra es más probable para ser considerada como *output*, como se vio en la figura 3 (3Blue1Brown, “*Transformers (how LLMs work) explained visually | DL5*” 20:30).

Así, pues, el *embedding* vectorial resulta una herramienta conceptual y epistemológica interesante y útil, bastante intuitiva para el cerebro humano y bastante refinada desde la perspectiva de las matemáticas. Como ya se ha señalado antes en 1.1, y en la cita de Jurafsky y Martin, esta es la forma en que los LLM tienen para comprender el significado de las palabras, de los textos, y, de cierta manera, del mundo: es la piedra angular de su epistemología. Luego esto se va, como ya también dijimos, refinando, los *weights* se van ajustando, se va filtrando cierta información en los *feedbacks*, y con ello tenemos un panorama bastante completo de la epistemología en las IA tipo LLM. De modo que para este caso no parece haber objeción alguna de que las IA no sean capaces de poseer competencias semánticas y pragmáticas. Ahora bien, el hecho de que se diga que esta disposición de palabras parametrizadas dispuestas en un espacio relacional es lo que permite que



las IA puedan entender el significado de esas palabras, y *tokens* en general, tiene su respaldo en otra gran base teórica de la lingüística: la hipótesis distribucional.

2.2.2 *El problema subyacente: la hipótesis distribucional y las competencias semánticas y pragmáticas de los LLM*

En esta base teórica, que da pie a todo el proceso que acabamos de describir, y al funcionamiento en general de los LLM, vemos la inevitable causa de las obras cerradas como algo inherente a estos modelos de IA. Volvamos a la postulación de las competencias semánticas y pragmáticas de las IA. Acordamos que el *embedding* vectorial era su manera de adquirir estas dos competencias. No sería de extrañarse, ya que como se señaló, dentro de la NLP, estas dimensiones lingüísticas como modelos epistemológicos se exploran a fondo. El punto de partida es que la comprensión de una palabra en un texto puede darse *gracias a la relación que guarda con otras palabras*. A este modelo semántico-contextual se le conoce mejor como *hipótesis distribucional*, y es otra de las teorías operativas de los Grandes Modelos de Lenguaje, el fundamento detrás de la operatividad de los *embeddings* vectoriales en particular, como lo validan Jurafsky y Martin:

Las palabras que aparecen en contextos similares tienden a tener significados similares.

Esta relación entre la similitud en la distribución de las palabras y su significado se denomina hipótesis distribucional. Esta hipótesis fue formulada por primera vez en la década de 1950 por lingüistas como Joos (1950), Harris (1954) y Firth (1957), quienes observaron que palabras sinónimas (como oculista y oftalmólogo) tendían a aparecer en el mismo entorno (p. ej., cerca de palabras como ojo o examinado), y la diferencia de significado entre dos palabras “corresponde aproximadamente a la diferencia en sus entornos” (Harris, 1954, p. 157). En este capítulo, presentamos la semántica vectorial, que



ejemplifica esta hipótesis lingüística mediante el aprendizaje de representaciones del significado de las palabras, llamadas *embeddings*...¹¹⁵ (101).

Luego, es de observarse que a la hipótesis distribucional (HD) en el contexto de los LLM se le otorgará la misma visión y función operativa respaldada por el determinismo numérico y la lógica matemática, bases a su vez de los LLM. Es decir: el significado de las palabras (semántica), y su relación entre ellas para dotarlas de contexto para comprender mejor su uso (semántica y pragmática) estará dado por números.

Es sobre esta noción, la hipótesis distributiva, en la que se basa gran parte del trabajo semántico en PNL, en particular en lo que se refiere a la vectorización de texto y las incrustaciones de palabras ... La idea es que si podemos asociar con una obra determinada ciertas métricas que se correlacionan con la distribución de la palabra a lo largo de un texto o corpus determinado, las palabras que tienen significados similares deberían a su vez tener propiedades de distribución similares en todo ese corpus textual determinado.¹¹⁶ (Worth, “Word Embeddings and Semantic Spaces in Natural Language Processing” 7-8).

¹¹⁵ Original: *Words that occur in similar contexts tend to have similar meanings. This link between similarity in how words are distributed and similarity in what they mean is called the distributional hypothesis. The hypothesis was first formulated in the 1950s by linguists like Joos (1950), Harris (1954), and Firth (1957), who noticed that words which are synonyms (like oculist and eye-doctor) tended to occur in the same environment (e.g., near words like eye or examined) with the amount of meaning difference between two words “corresponding roughly to the amount of difference in their environments” (Harris, 1954, p. 157). In this chapter we introduce vector semantics, which instantiates this linguistic hypothesis by learning representations of the meaning of words, called embeddings...*

¹¹⁶ Original: *One of the fundamental tenets of linguistics that has been well established since the middle of the last century is the idea that the meaning of a word can be understood by not just the context within which it is found (or heard as the case may be) but that generally, over a large corpus of text, or a large sampling of language as the case may be, that words found within a similar context should in most cases have similar meanings, leading to the now oft-used and famed expression in both linguistics and NLP circles that a word can be characterized, or understood in some sense, by the company it keeps. This is known in linguistics as well as NLP circles as the distributional hypothesis and it is this tenet that underpins many of the techniques described here which fall under the heading of word embeddings. It's upon this notion, the distributional hypothesis, that much of the semantic work in NLP is based, in particular as it relates to text vectorization and word embeddings, the topic of this paper. The idea being that if we can associate with a given work certain metrics that correlate to the word's distribution throughout a given text or corpus, words that have similar meanings should in turn have similar distributional properties throughout that given textual corpus.*



Comprender un término para un LLM implica comprenderlo en función de los números que componen a ese concepto o término. Mientras que para el ser humano, la comprensión de un término es una labor semiótica, o bien lingüística, en términos de asignar a un significante un significado, de asignar a un signo un interpretante y un objeto, procesos en que intervienen no pocas funciones cognitivas¹¹⁷, como las memorias a largo y corto plazo, por ejemplo, además de otros tantos cuadros sociales y contextuales¹¹⁸. Así que si se puede afirmar que estas tecnologías realmente *comprenden* los datos que manejan, ello debe entenderse como que los comprenden dada la lógica matemática y no una comprensión similar a la humana, obviamente. La lógica matemática continúa siendo el filtro principal e inamovible con el que opera el mecanismo de los *embeddings*, filtro inherente a toda aparato epistemológico de las IA tipo LLM:

este enfoque se basa en un conjunto de teorías lingüísticas que, a su vez, se formulan matemáticamente o, más precisamente, estadísticamente, en un área de investigación y desarrollo que a veces se denomina *semántica distributiva*, donde las similitudes entre términos lingüísticos se derivan de las propiedades distributivas de los términos subyacentes. En este contexto, los *embeddings* de palabras pueden entenderse como representaciones semánticas de una palabra o término determinado en un corpus textual determinado. En esencia, los *embeddings* de palabras pueden entenderse como un conjunto de números, atributos o características (reales) generados matemáticamente (estadísticamente) dentro del contexto del *Machine Learning*, que se crean a partir de un cuerpo de texto que asigna una palabra determinada a un espacio semántico para un

¹¹⁷ “El significado semiótico está ligado al significado cognoscitivo” (Eco, *Semiótica y filosofía del lenguaje* 130). Ahora bien, véase la extensa nota 133 donde se introduce una discusión del mismo Eco —y Peirce— sobre esta perspectiva, que la pondría en jaque.

¹¹⁸ Este recorrido teórico es según la teoría de Teun A. van Dijk, sobre todo como está expuesta en su obra *Estructuras y funciones del discurso*.



dominio determinado que se desprende de la hipótesis de distribución¹¹⁹ (Worth, “*Word Embeddings and Semantic Spaces in Natural Language Processing*” 11).

El argumento que extendemos a esta propuesta no va tanto en función de criticar que dicha forma de significar, de entender, tiene sus fallas frente a otro tipo de modelos epistemológicos, sino que *esta visión es restrictiva para las áreas creativas*. Sin embargo, para poder formular un argumento mejor cimentado en cuanto a esta visión restringida para la creatividad, área celosamente resguardada para el ser humano —y quizá para algunos animales o la naturaleza—, habría que ahondar algo en esa crítica que va directo a la pertinencia de la hipótesis distribucional. De entrada, porque nos parece conflictiva la aproximación que toma la HD, dado que entendemos que la creatividad que demandan las artes, el pensamiento en general que deviene en las Humanidades, no reposa sus bases en las relaciones numéricas arbitrarias que propone la HD, sino en conceptos más propiamente humanos, aun si se trata en términos de, para unos, conceptos materiales. Extenderemos, pues, algunas críticas a la validez de la HD en función de su uso en el contexto de la creación literaria.

Una de estas críticas es, por ejemplo, que el concepto de similitud semántica es demasiado extenso y difícil de capturar y distinguir en un modelo distribucional, a lo que antepone el argumento de que existe una noción amplia de similitud semántica (Shalgren, “*The distributional hypothesis*” 37). Por otro lado, expone también un punto a favor de la HD: que esta podría tener

¹¹⁹ Original: *This approach however relies on a certain (set of) linguistic theories which then in turn are formulated mathematically, or more precisely statistically, in an area of research and development that is sometimes referred to as distributional semantics, where similarities between linguistic terms are derived from the distributional properties of the underlying terms. In this context, word embeddings can be understood as semantic representations of a given word or term in a given textual corpus. At their core then, word embeddings can be understood as a set of mathematically (statistically) generated (real) numbers, attributes or features within the context of ML, that are created from a body of text which map a given word into a semantic space for a given domain which follows from the distribution hypothesis.*



un valor y un enfoque estructuralistas, es decir, que su validez es únicamente cuando se la aplica dentro de un texto.

La crítica común que se hace a los métodos distributivos es que no son suficientes para llegar a una semántica completa porque sólo analizan los patrones distributivos del texto. Como se ha demostrado en este artículo, esa crítica es irrelevante porque los únicos significados que existen dentro de una explicación estructuralista del lenguaje son los tipos de relaciones que adquieren los métodos distributivos.

Los modelos distributivos son modelos del significado de las palabras. No de los significados que están en nuestras cabezas ni de los significados que están en el mundo, sino de los significados que están en el texto. El paradigma distributivo puede ser un caso extraño, pero constituye, no obstante, una forma viable de llegar al significado (49).

El argumento es válido cuando dice que es una forma viable de llegar al significado: al menos así lo es para los LLM, que trabajan bien sobre esta base¹²⁰. Pero las relaciones que se establecen con los LLM, al menos en el caso de los chatbots, no se trata de relaciones meramente de lecturas de texto, por lo que esta visión del paradigma distributivo nos parece que levanta todavía demasiadas limitaciones en su forma de aproximarse al significado de las palabras, y aun a su tratamiento como actos pragmáticos, como actos de habla.

Otra de las caras de las críticas a la hipótesis distribucional se constituye precisamente por señalar que su uso está en función y a la orden de la programación y en áreas de la ciencia computacional, bases del aprendizaje de máquinas. El enfoque de las siguientes observaciones es el de que la semántica distribucional, “...sólo puede permitir inferir el uso de una palabra, pero no

¹²⁰ Véase también este respaldo de la pertinencia de este tratamiento del significado de las palabras dentro del texto para los LLM en Jurafsky y Martin, Speech and Language Processing: An Introduction... 106-111.



su significado¹²¹” (Ottolenghi, Stefano, “*The Distributional Hypothesis: semantic models in theory and practice*”). Esta distinción es la que hemos comentado previamente.

Una crítica fundamental a la hipótesis distributiva es que, en la medida de lo posible, los modelos semánticos distributivos no pueden ir más allá del aprendizaje del uso de las palabras. De hecho, aprender el significado de las palabras queda totalmente fuera de su alcance¹²² (Ottolenghi, Stefano, “*The Distributional Hypothesis: semantic models in theory and practice*”).

En este punto, el autor de artículo aborda el ejemplo expuesto por John Searle, una suerte de reinterpretación del juego propuesto por Turing, que tituló *Chinese Room*, para respaldar la idea de cómo a través del enfoque de la HD, lo que se hace, más bien, es inferir un cierto uso de las palabras, pero sin apoyarse realmente en competencias semánticas y pragmáticas sólidas. Y lo que es más: el argumento del cuarto chino es todavía más verificable para los LLM, tal como lo sería el símil del autómata turco que jugaba ajedrez. La idea general de Searle, y que aterriza en las observaciones sobre las competencias semánticas y pragmáticas, es la siguiente:

Una persona que no sabe chino, como yo, por ejemplo, está encerrada en una habitación y, como dice Turing, está equipada únicamente con cosas como un lápiz y un trozo de papel, junto con otros elementos que forman la base de datos y el programa, en este caso símbolos chinos, y un programa de instrucciones sobre qué hacer con los símbolos chinos. En el experimento mental, la persona que está en la habitación, es decir, yo, es capaz de simular el comportamiento de una persona china de una manera que satisface el Test de Turing, porque, por ejemplo, da las respuestas correctas a las preguntas en chino, pero de todos

¹²¹ Original: ...may only allow to infer the usage of a word, but not its meaning.

¹²² Original: One fundamental critique to the distributional hypothesis is that, as far as it could go, distributional semantic models may not go beyond learning word usage. In fact, learning words meaning is totally out of their scope.



modos no entiende una palabra de chino. ¿Por qué no? Porque no tiene forma de llegar desde la manipulación de los símbolos a los significados de los símbolos, y si la persona que está en la habitación no entiende los significados de los símbolos sobre la base de la implementación del programa, entonces tampoco lo entiende ningún otro ordenador únicamente sobre la base de eso, porque ningún ordenador, simplemente en virtud de sus propiedades computacionales, tiene algo que el hombre no tenga. Además, toda la habitación no tiene forma de llegar desde los símbolos a sus significados. El sistema en su conjunto no tiene forma de asociar una semántica a la sintaxis de los símbolos informáticos (“*The Turing Test: 55 Years Later*” 144).

Adela Cortina, a quien recurriremos más ampliamente en el capítulo 3, respalda la visión de Searle y añade otros criterios a esta. Searle fue también quien discutió las posibles variantes de IA fuerte y débil. Según lo que postula en el ejemplo del *Chinese Room*, una IA que realmente pueda entender la semántica del juego, es una IA fuerte. A esto hace la siguiente revisión la filósofa:

... lo que intenta demostrar Searle es que la IA fuerte es imposible, porque la máquina carece de la *intencionalidad* por la que los humanos damos significado a lo que nos rodea: una máquina no conoce el significado de los símbolos que maneja. *Sin un cuerpo*, las representaciones abstractas carecen de contenido semántico: no puede haber inteligencia general sin un cuerpo ... las máquinas carecen del conocimiento del sentido común que es posible por nuestras vivencias corporales. El cuerpo es esencial para dar significado a lo que nos rodea mediante la intencionalidad (*¿Ética o ideología de la inteligencia artificial?* cap. 2)

Intencionalidad, corporalidad y, por ende, sentido común, son la clave de este argumento de Cortina, observado desde Searle, que hace frente a los argumentos que validan la HD, señalando



que es funcional porque *se aproxima* a lo que es la comprensión de las cosas, o porque tiene *cierta validez* desde cierto punto de vista. Por otro lado, la discusión obviamente no se zanja aquí, la misma filósofa admitirá más adelante en este pasaje que no es imposible que una forma de sentido común se pueda lograr codificar para las IA, hecho que, por supuesto, hasta no ver cómo se desarrolla no podríamos examinar, pero desde ya, con lo que hemos expuesto, podemos prever que no estará exento de críticas¹²³.

La discusión que ha derivado de este apartado hasta este ejemplo de Searle, tiene una fuerte base también en la filosofía del lenguaje, como se verá a continuación. Hemos supuesto, como Searle y Cortina, que *el mero uso de las palabras, sin intencionalidad (y por tanto, sin competencia pragmática, necesaria para la comunicación y aun para reafirmar la semántica de los términos), no constituye la totalidad de la comprensión semántica de los términos*. Y esto nos deja nada más y nada menos que a las puertas de la teoría de Wittgenstein, sobre todo del llamado “segundo Wittgenstein”.

El postulado principal de la filosofía del lenguaje del segundo Wittgenstein es el de que el significado de una palabra está dado por su uso (*Investigaciones filosóficas* 171, 191, por ejemplo). Este postulado, que abre y permea toda su obra subsiguiente, es la base de lo que se conoce como “juegos del lenguaje”. Bajo este criterio, una IAG definitivamente conoce el significado de las palabras, porque no se puede negar que sabe usarlas¹²⁴. Pero la cuestión, una vez más, no es

¹²³ Parece ser que la teoría del cuerpo, a la que también apunta Cortina, sería un buen punto de anclaje para la formulación de un aparato crítico que sopesa la validez de cualquier demostración de capacidad cognitiva y aun emocional de una IA.

¹²⁴ Sobre lo dicho de Wittgenstein desde la PLN y cómo este representa una base teórica que respalda la utilidad de la HD para los LLM: “Parte del fundamento filosófico del modo de pensar distributivo proviene de los últimos escritos del filósofo Wittgenstein … La idea de Wittgenstein es que deberíamos definir una palabra por cómo la usan las personas al hablar y comprender en sus interacciones cotidianas, prefigurando así el movimiento hacia modelos encarnados y experienciales en la lingüística y la PNL…” (Jurafsky y Martin, *Speech and Language Processing: An Introduction…* 130) (Original: *Some of the philosophical underpinning of the distributional way of thinking came from the late writings of the philosopher Wittgenstein ... Wittgenstein's idea is that we should define a word by how*



sencilla: cuando se trata del lenguaje, casi nada tiende a serlo. De entrada, pese a lo dicho, en Wittgenstein también existe un choque relevante con esta visión de significados estrictos para los términos:

El punto de Wittgenstein es que las palabras y los lenguajes no pueden ser reducidos a esencias o definiciones estrictas. El significado de una palabra depende, en última instancia, de la forma en que es usada en un contexto específico, y cuando uno sabe cómo usar una palabra entonces uno conoce su significado. Sin embargo, saber cómo usar una palabra no significa tener una definición de la misma (Robinson, “Wittgenstein, sobre el lenguaje” (30).

De manera que volvemos a la idea de que el uso de una palabra debe ir acompañado por cierta intencionalidad¹²⁵. El hecho de que los LLM no acompañen sus textos, discursos y palabras con intencionalidad no invalida sus palabras al grado de que todas ellas deberían considerarse como patrañas¹²⁶, pero, argumentamos, eso las coloca de nuevo en la desventaja de no comprender realmente el significado de las palabras. Y ya desde que esto queda establecido así, son muchas las consecuencias que se desprenden¹²⁷, empezando por una separación evidentemente marcada con la inteligencia y las capacidades humanas, algo que invitaría a replantear la cita de Elaine Rich, cuando menos el enfoque de la palabra inteligencia en el concepto de inteligencia artificial.

Para dar más claridad a la discusión, que es bastante densa, acudimos a un ensayo de Eco dentro de *Semiotica y filosofia del lenguaje* que confronta los modelos semánticos de diccionario

it is used by people in speaking and understanding in their day-to-day interactions, thus prefiguring the movement toward embodied and experiential models in linguistics and NLP).

¹²⁵ Más tarde en su obra, el segundo Wittgenstein insistirá de nuevo en los usos del lenguaje, pero llegando a estos desde el pensamiento, desde procesos mentales, lo cual añade otra capa al uso del lenguaje (Katz, La realidad subyacente del lenguaje y su valor filosófico 22).

¹²⁶ Esta discusión se abre de manera amplia en 3.2.

¹²⁷ Posibilidades de exhibir sesgos, *bullshit*, alucinaciones, replicar visiones hegemónicas del mundo y, por supuesto, crear obras cerradas en lo que a creación literaria se refiere.



y enciclopedia utilizados dentro de la semiótica, confrontación que termina por abordar a Wittgenstein y el significado. Eco expone la importancia de la pragmática para completar el significado, a lo cual termina por abordar a Wittgenstein y su argumento de las *Investigaciones*, señalando luego una contradicción en la que incurre el filósofo: “poco después el propio Wittgenstein se contradice al reconocer que ‘la explicación del significado explica el uso de la palabra’ ...” (91). Esta revisión del postulado concuerda con el final de la cita de Robinson, puesto que para el uso de una palabra primero debe conocerse su significado, incluso cuando se trata de un significado aprehendido ostensivamente, a la manera incluso de San Agustín¹²⁸. Esta aproximación a Wittgenstein le sirve a Eco para luego proponer un modelo de *semántica componencial* que se postule como un mejor modelo para la significación, modelo que prevería las relaciones semánticas y pragmáticas de las palabras en los textos. La semántica componencial incluye de cierta manera una porción de una semántica distribucional, puesto que se interesa por las relaciones entre términos, pero no remite a que el significado de estos resida meramente en ello. La base que propone Eco para un análisis componencial es la enciclopedia, términos que ya hemos abordado antes. En un despliegue de enciclopedia, se pueden ver tanto las relaciones entre términos, que operarían también de una manera distribucional, pero también el proceso de semiosis, donde un signo es interpretado por una larga cadena de interpretantes, es decir, un signo de un signo. Esta cadena alcanza a cubrir también usos contextuales de los signos, por ejemplo.

La fecundidad de la noción de interpretante no deriva sólo del hecho de que describe la única manera en que los seres humanos establecen, acuerdan y reconocen los significados de los signos que utilizan. Esa noción es fecunda porque muestra de qué manera los procesos semióticos ... circunscriben asintóticamente los significados (o los contenidos, en

¹²⁸ Beuchot, *Historia de la filosofía del lenguaje* 51.



una palabra las ‘unidades’ que la cultura ha determinado en el curso de su proceso de asignación de pertinencia al contenido), sin llegar nunca a tocarlos directamente, pero volviéndolos de hecho accesibles mediante otras unidades culturales. Esta continua circularidad es la condición normal de los sistemas de significación y se lleva a la práctica en los procesos de comunicación (131).

Un modelo de esta índole expone con mucha mejor representación un modelo de significado, que se basa en abrir recorridos, casi infinitos, de interpretación de los términos. Lo que en la semántica vectorial serían relaciones de antonimia u oposición, también estarían contemplados en el modelo de la enciclopedia. Pero la calidad rizomática de esta, y el hecho de que contemple también una pragmática de las palabras, empujando a procesos de comunicación, la instala como un modelo más cercano a los procesos cognitivos del ser humano, que implican el ejercer interpretaciones¹²⁹. Una observación que vemos como evidencia de que este modelo podría ser más fructífero como modelo semántico, es que a través del modelo de enciclopedia que delinea la semántica composicional, los sesgos, por ejemplo, se podrían manejar con más cautela, codificarse históricamente y en sus varias formas y acepciones¹³⁰. En esta cuestión, todavía hay mucho que criticarle a la semántica distribucional, que no logra manejar los sesgos de un modo óptimo¹³¹.

¹²⁹ Diría Peirce sobre este proceso: “Un signo o representamen es algo que está para alguien en lugar de algo en cuanto a algún aspecto o capacidad. Se dirige a alguien, es decir, crea en la mente de esa persona un signo equivalente o, quizás, un signo más desarrollado” (citado desde Eco, *Lector in fabula* 42). Esta cadena infinita es la que se postula con el concepto de enciclopedia, propio de una semántica composicional, resultado de la puesta a prueba de otros modelos semánticos previos (árbol de Profirio, Modelo Q, Árbol KF).

¹³⁰ Como ya se mencionó anteriormente, en otro lugar de su obra semiótica, Eco parece optar o intercambiar el modelo composicional por el instruccional, el cual es, de hecho, posible de codificar para el entendimiento de un ordenador y aun de una inteligencia artificial, si bien no de una del tipo Transformador, por supuesto (*Los límites de la interpretación* 247-248). Esta discusión se amplía en la nota 131.

¹³¹ Véase en Jurafsky y Martin, *Speech and Language Processing: An Introduction...* 126, los intentos por codificar una semántica histórica, que tenga en cuenta empleos descartables del lenguaje en la actualidad, y en 127 cómo este y otros métodos propuestos siguen sin poder tratar adecuadamente el problema de los sesgos.



El de Eco no es, por supuesto, el único modelo semántico, como tampoco el de la HD y estas no son las únicas teorías que se debaten sobre la calidad del lenguaje¹³². El señalar esto no nos remite sino a la idea de que el problema del significado es uno amplio, que no se resuelve ni en uno ni en otro modelo. No obstante, al inicio de este subapartado establecimos que no era nuestra labor evidenciar lo falible de la HD *per se*: lo es en cuanto que nos demuestre que es una causa directa para la generación de obras cerradas y de estéticas constreñidas. Podemos continuar la discusión por este derrotero a través de la exploración de la principal de las figuras retóricas, presente en el habla de la vida diaria, base misma del lenguaje: la metáfora.

La visión de significado que planteamos, donde intervienen varios factores tanto semánticos, textuales, cognitivos y sociales, es una visión que se tiene en cuenta en la amplitud de procesos que compone a la metáfora. En *Semiotica y filosofia del lenguaje*, Eco continúa con la discusión sobre el significado. La visión de la enciclopedia favorece notablemente la producción de metáforas ricas en contenido. Una metáfora, desde la semántica composicional, es el intercambio de un semema por uno de sus semas. La exposición de Eco en esta sección concluye que la riqueza de la metáfora está en el recorrido semiótico que traza la metáfora, es decir, en la cadena de interpretaciones que se pueden asignar a un signo. Un término, es decir, un semema, despliega una cadena, mejor dicho, una ramificación, o rizoma, de semas, de partes que dotan de significado, y que sirven aun para interpretar ese sema. La metáfora se da cuando el semema es

¹³² Como referencia, véase a Katz, que como parte de la escuela de la gramática generativa transformacional propone una dicotomía entre las concepciones de lenguaje de diferentes escuelas: las que consideran al lenguaje como de calidad democrática y las no-democráticas. La teoría democrática sostiene que las reglas gramaticales del lenguaje verbal reflejan las estructuras profundas del lenguaje, reglas mucho más abstractas y complicadas de organización lingüística, que ponen de manifiesto que la estructura latente es mucho más profunda que cualquier otro tipo de estructura superficial (29). Incluso en este tratamiento del lenguaje, la HD queda comprometida en cuanto que su forma de crear significado remitiría a una visión no-democrática, ya que no alude a ningún tipo de estructura profunda: lo mismo que en Wittgenstein, su búsqueda del significado acaba por relaciones externas con otras palabras. Véase en 129 una crítica a la aproximación de las gramáticas taxonómicas dentro de las cuales se ubican diversos métodos distributivos de clasificación del lenguaje. Una vez más, subrayamos que esto es solo otra disyuntiva en el problema, una de tantas como tantas son las escuelas lingüísticas que las plantean.



reemplazado por uno de sus semas. En ese momento hay ya un recorrido semiótico a lo largo del rizoma. La visión de esta figura, de este despliegue enciclopédico, para visualizar los recorridos de una metáfora tiene, por supuesto, un centro, o un comienzo, y una periferia. La idea principal es que entre más alejado sea el sema que se intercambiará por el semema central, más rica resultará la metáfora. Un ojo avisado, como el del poeta, establecería una conexión semántica no del todo contemplada, es decir, conectaría un sema más o menos periférico al semema en cuestión, logrando un efecto incluso novedoso en apariencia. Esta comprensión del proceso puede trasladarse a proposiciones más grandes, a textos completos, por ejemplo, y así tener un atisbo de los mecanismos que producen textos interesantes y creativos, es decir, que incursionen en conexiones no antes establecidas, en este caso conexiones semánticas, pero también podrían serlo pragmáticas.

Así, pues, concreta Eco:

El modelo de enciclopedia construido para llevar a cabo la interpretación de determinado contexto es el que fija *ad hoc* dónde está el centro y dónde la periferia de los semas. Subsiste el criterio de la mayor o menor apertura, es decir, de la extensión del viaje que la metáfora permite realizar a través de la semiosis, y del grado en que permite conocer los laberintos de la enciclopedia. En el curso de ese viaje, los términos que intervienen adquieren nuevas prioridades que la enciclopedia aún no les había reconocido (224).

Y si bien estos criterios no bastan para establecer una estética de la metáfora, para poder hablar de metáforas “feas” y metáforas “bellas”, “*bastan para distinguir entre la metáfora cerrada (o poco cognoscitiva) y la abierta, que permite conocer mejor las posibilidades de la semiosis*” (225). He aquí la clave, un criterio clave observable en las inteligencias artificiales generativas.

Piénsese de nuevo en los modelos de los *embeddings* y su operatividad amparada en el concepto de la hipótesis distribucional. Como se vio en las figuras 5 y 6, los términos tienen una



orientación y ello determina un cuadro contextual para ese término y sus términos allegados. Y esto es un mecanismo, de hecho, bastante ideal desde la perspectiva estructural del texto, como ya se dijo. *Pero no lo es desde la perspectiva de la literatura.* Los *embeddings* están “fríamente” calculados, algebraicamente dispuestos, de manera que todo encaja en la lógica matemática. Pero, como hemos observado con el ejemplo de la metáfora, la literatura exige *el rompimiento de moldes como estos*. Si un poeta escribiera con el orden y la lógica de los *embeddings* no estaría haciendo más que adoptar un campo semántico muy firme y ordenado para su texto. Pero la belleza de las letras, lo sorprendente de las tramas y sus clímax, la belleza de una prosa, lo atractivo de un estilo, etc., *no provienen de operar con esos mecanismos con los que operan las máquinas*, ¡todo lo contrario! “No hay un algoritmo para la metáfora: no puede ser prescrita a un ordenador mediante instrucciones precisas, independientemente del volumen de información organizada que podamos proporcionarle” (225). Si bien Eco, dadas sus bases teóricas, admite que el intérprete, o lector de la metáfora, con toda su carga ideológica y contextual, administra una buena parte del sentido a la metáfora, está claro que la metáfora tiene también esta disposición de abierta o cerrada para esa interpretación de su lector, y entre más riqueza tenga esta en su recorrido semiótico, mejor será su conexión con la interpretación del lector, hecho pragmático y neurológico a la vez (228), factores que crean una separación para lo creativo entre la mente humana y la operatividad de la máquina¹³³.

¹³³ Únicamente con afán de abrir el camino a una discusión futura —y de admitir que la problemática del presente apartado es demasiado amplia y que no nos podemos adjudicar la última palabra en unas cuantas líneas—, aludiremos aquí un caso que podría parecer controversial, cuando no contradictorio. El mismo Eco parece haber mantenido la convicción de que un modelo semántico perfeccionado, es decir, su enciclopedia idealizada, operando sobre la base de un modelo semántico instruccional, podría ser alimentado a una inteligencia artificial. Una vez en poder de este mecanismo epistémico, la forma de operar de la IA en cuestión sería, de alguna manera, *igual* a la del pensamiento humano, o cuando menos sin diferencias discutibles. La idea se esbozaba ya desde *Signo*, en donde se discute la posibilidad de una teoría unificada del signo, que años más tarde desembocará en su concepto de enciclopedia, y se concluye que si bien esta hipotética mente artificial, capaz de manejar esta teoría unificada, “no es enteramente igual a un hombre, con todo, se aproximaría bastante a lo que hay en el hombre de comportamiento signíco socializado” (168). Pero más concreta se manifiesta en *Los límites de la interpretación*, en el ejercicio imaginativo del apartado 4.5: “Charles Sanders Peirce”, cuyo mensaje es interpretable como la posibilidad de que si una IA pudiera adoptar un modelo de semántica instruccional, o enciclopedia, su pensamiento no sería muy distinto de aquel de un ser humano, puesto que habría dominado, de cierta manera, todos los saberes posibles del mundo, interconectándolos,



2.2.3 El Autor Modelo y la obra cerrada como resultados del determinismo numérico del embedding vectorial

En síntesis, precisamente a través de esta visión del significado que exploramos en este subapartado y su conexión con la metáfora, es que podemos entender de mejor forma lo que aseveramos al final del apartado 2.1: el mecanismo de *embeddings* vectoriales tiene implicaciones importantes para la creación literaria de las IAG. Más puntualmente: la epistemología que construye el mecanismo de los *embeddings* vectoriales restringe la obtención de significado de los LLM bajo la lógica matemática y el determinismo numérico, factores útiles a la hora de realizar cálculos y tareas sistemáticas, pero no al momento de la creatividad, de la búsqueda de las metáforas abiertas, de las obras abiertas. De modo que la obra cerrada, reiteramos, es inherente a las IAG. Pero este enunciado lleva una aclaración familiar: la obra cerrada es inherente a las IAG cuando se les usa para la creación literaria en niveles bajos de intervención del usuario.

utilizándolos acorde a contextos y circunstancias adecuados (355-356), posibilidad que incluso ayudaría a codificar aun mejor los ya mencionados sesgos y patrañas, sobre todo, como ya se dijo con respecto a la semántica distribucional. No obstante la lucidez de la propuesta de Eco, debe leerse siempre desde las perspectivas adecuadas: no sólo es la propuesta un eco lejano, si bien más sensato, de las neurosis de los gramáticos renacentistas como Wilkins y Dee, pero sobre todo del filósofo, místico y teólogo medieval Raymundo Lulio, que los impulsó a buscar la codificación de todo el saber del mundo en modelos finitos, pero de combinatorias quasi infinitas. La enciclopedia, acarreada por la semántica instruccional, no necesariamente pretende prever todo el saber del mundo a la manera de estos personajes, pero sin duda cae en un reduccionismo si se usa para sustentar que puede hacer cognitivamente indiferenciable a un ordenador dotado de IA de un ser humano pensante. Quizá es por eso que Eco no hace la afirmación directamente, ni en *Signo*, ni en *Los límites*, sobre todo en este, donde lo postula a través de un ejercicio ficcional, si bien bastante cimentado, hasta por citas de la obra de Peirce. Mucho tiene que ver la postura de Eco ante el problema semiótico del referente, visión que comparte con Peirce, por el cual no hace falta la presencia de una mente humana para que exista la semiosis. Es decir, Eco no comparte la visión de una semiótica mentalista, lo cual es abrir un terreno seguro para el argumento de la no diferenciabilidad que subyace al ejercicio “Charles Sanders Personal”, y aun para toda su teoría semiótica. Más allá de esto, se debe contextualizar esta creencia en una época en que los experimentos lingüísticos con las IA profesaban no de poco optimismo, pero se debe asimismo recordar que los modelos de los años 80-90s eran muy distintos a los de hoy, y todavía permanecía demasiado avivado el influjo de la teoría de Turing y su máquina pensante.

Así, pues, creemos que esta visión de Eco no invalida las propuestas y argumentos hasta este punto erigidos frente a la semántica distribucional y el determinismo numérico como base teórica de la IA. En todo caso, la admitimos como otra forma de abordar futuras discusiones sobre la verdadera comprensión del significado en estas tecnologías, cuando no sobre la posibilidad de toda una semiótica de la inteligencia artificial.



Todavía podemos comprobar este fenómeno “en la práctica” volviendo a una discusión que dejamos truncada en 1.3 y que completaría la noción del Autor Modelo.

Ahora que tenemos conocimiento de esta cerrazón de obras textuales que son los LLM, podemos decir que *el Autor Modelo de ChatGPT4 está condicionado fuertemente por el mecanismo del embedding vectorial y su tratamiento de la semántica y la pragmática de los términos*. Y esto nos lleva a otra conclusión también: *dado que un AM depende también de la intervención del usuario (IU), una baja IU determinará, por tanto, una mayor libertad de operación al AM generado por ChatGPT4 y, por ende, un texto creativo bastante pobre en sus cualidades y posibilidades, en sus formas y contenidos*.

Los *embeddings* se agrupan muy ordenadamente por sus cualidades semánticas atribuidas desde la perspectiva distribucional. De modo que el prompt del usuario será interpretado desde este criterio. Cuando hemos sugerido un cuento que contuviera sememas como “fantasmal”, “policíaco”, “detective”, la lectura de ChatGPT4, con base a *embeddings* vectoriales, comenzará una labor que atraerá sólo a los *embeddings* más próximos a estos sememas. Como hemos dicho antes: será un proceso bastante ordenado desde el punto de vista semántico, pero jamás causará demasiada sorpresa porque no habrá un registro contrastante de sememas. Lo |fantasmal| será únicamente asociado con lo fantasmal y ya, como un imán que atrae los detritos de grafito próximos a él nada más; |detective| será asociado con sus *embeddings* cercanos sin posibilidades de introducir un elemento lejano, periférico, que agregue todavía un toque de mayor incertidumbre, un extrañamiento deliberado que amplíe las posibilidades del texto. Todo esto, que está en las capacidades de cualquier escritor humano mínimamente competente, sólo pueden contemplarse en una IA como errores, a fin de cuentas.



Lo habíamos mencionado ya en el estado del arte crítico con el estudio de Inna Adamivna, “*The Art of Narration and Artificial Narrative Intelligence: Implications for Interdisciplinary Research*”: “las teorías existentes sobre la narrativa en IA reflejan una comprensión de la representación basada únicamente en construcciones textuales, lógicas o matemáticas¹³⁴” (315), de ahí que la investigadora sugiera nuevas aproximaciones a la exploración de las IA en sus modelos narrativos y modelos de comprensión de la literatura, algo que, ahora lo sabemos, es sumamente complicado, ya que requeriría un cambio a nivel de las bases conceptuales de los LLM.

Este hallazgo sobre los LLM presenta muchos matices. Por un lado, insistimos en lo fructífero que puede resultar dentro de una semántica textual, o en sí como una aproximación de núcleo estructuralista, para la lectura de textos. Hemos dicho que para cualquier labor de índole algebráica, los resultados serían muy buenos. Incluso en el reconocimiento de imágenes habría más posibilidades, quizá. Pero los fallos que presenta hacia la creación literaria hacen de este enfoque uno nada favorable para una poética contemporánea de literatura de obras abiertas. El enfoque mismo de reducir el conocimiento a números, lo cual sustenta el determinismo a que están sujetos los Transformadores con sus redes neuronales, y por tanto las IAG en general, remite de buen grado a la concepción pitagórica del mundo, pero quizá sin todo el revestimiento platónico de la época Clásica, o del neoplatonismo renacentista incluso. Más bien opera con el revestimiento de un materialismo científico-matemático, o positivista, lo cual no puede significar otra cosa que un reduccionismo en la comprensión del significado de las cosas, hecho que, como se vio, remite a restricciones epistemológicamente importantes en otras áreas, como la literatura y el arte.

En este punto, esta discusión de una generación de obras siempre cerradas conduce también a una discusión sobre la experiencia estética del arte, y aun a una discusión sobre los códigos

¹³⁴ Original: *existing theories of narrative in AI reflect understanding of representation based only on textual, logical, or mathematical constructs.*



estéticos del arte en la cultura. La IAG como creadora literaria, o como herramienta para la literatura, quizá no esté del todo codificada, aceptada en la cultura para interpretarse con este uso, pero ya vimos que no es un uso que le sea ajeno. Con todo y los problemas que presenta se han abierto espacio dentro de la cultura. ¿Por qué?

Si lo que queremos es terminar de entender la cooperación interpretativa en los “textos artificiales”, lo que resta por hacer es dar una mirada al camino que esta propuesta estética plantea. Hay material suficiente para dar pie a estas reflexiones. Incluso lo que hemos entendido como errores, como fallos en el proceso de *embedding* y *unembedding*, o cualquier fallo general en los *outputs* de las IAG puede constituir material de reflexión estética. Luego, hemos dejado también incompleto el concepto del Lector Modelo en tanto que todavía es más lo que de este se puede decir si consideramos este siguiente apartado. Porque una consideración importante para la estética de los textos de las IAG, viene dada, de nuevo, por la cantidad de intervención del usuario. Y esta fue, de hecho, la conclusión a que nos condujo el apartado del Lector Modelo de ChatGPT, por lo que el encuentro entre estos dos temas es inminente. El nodo en que se unen es la cantidad de intervención del usuario, pero también la calidad de la intervención. Todos estos factores enunciados empiezan a constituir ya un esbozo de una teoría estética verificable en el uso de la IAG, donde la obra cerrada debe de ser considerada, por supuesto. En general es de considerarse una estética de las IAG porque, de infiltrarse cada vez más las inteligencias artificiales en la cultura, estas reflexiones se volverán también inminentes.

2.3 Criterios para una estética de las creaciones literarias de las IAG

La discusión previa desemboca en la reflexión estética en tanto que las estrategias textuales de AM y LM, y los criterios de obra cerrada y obra abierta, son también construcciones estéticas.



Ya hemos dicho que, si no estuviéramos hablando de máquinas, la discusión sería concerniente más bien a un tipo de crítica especializada, que efectuaría su crítica con base a su definición estética del texto artístico texto. Pero las IAG plantean sus propios problemas estéticos para la literatura¹³⁵. Los hallazgos del apartado anterior no hacen sino énfasis en ello: si los modelos de producción creativa de las IAG son eso, modelos, con cristalizados mecanismos deterministas y lógicos, entonces sus creaciones literarias pueden ser motivo de una nueva postulación estética, que las analice en función de su actuar en la cultura. Ahora bien, esta discusión puede tener un retroceso: si se considera que lo que las IA producen no es literatura, ya sea por la falta de facultades humanas, como las que hemos recuperado antes de Cortina, como sentido común y corporalidad, facultades, por cierto, no menores, entonces no tiene sentido plantear ni siquiera la idea de una estética. Pero no es la creencia por la que nos guiamos aquí. Precisamente porque si estamos hablando de libertad interpretativa en la literatura, esta visión debe estar amparada por una más amplia que prevenga de cualquier visión dictatorial sobre el arte. No por ello dejarían de existir críticas lo suficientemente lúcidas y respaldadas para afirmar que las inteligencias artificiales y la creación literaria son temas separados. Pero la tarea de la semiótica también es recoger los signos de la cultura, siendo esta un producto del intercambio de signos, es decir, de la comunicación. Y si hemos atestiguado la presencia de las IAG en la cultura, sobre todo en apartados como el planteamiento del problema, es válido entonces que hagamos en este punto una reflexión estética, y que aceptemos que, al menos en tanto que la labor creativa de las IA es en coautoría con un ser humano, estamos tratando con textos literarios.

¹³⁵ Lo ya dicho en Tena y Flores, “Criterios para la valoración de la creación literaria de las inteligencias artificiales generativas” 4-5.



No obstante, aun con esta postura receptiva nos apegamos también a la postura que determina que los aparatos críticos para las IAG deben ser más “refinados” que los normales. Esta idea subyace a toda esta investigación. De modo que el siguiente paso es plantear las consideraciones de una estética dirigida a las IAG y sus creaciones literarias. Hay importantes restricciones que se agregan a esta forma de pensar: no es posible plantear en este espacio más que una visión estética preliminar, pero sí podemos discutir los principios con los que podría operar una teoría más amplia. La primera restricción es que nos enfocaremos en el texto literario narrativo, considerando los hallazgos hasta ahora acumulados. No significa ello que las reflexiones no sean fructíferas, sin embargo: hemos observado ya ciertas tendencias en las inteligencias artificiales generativas en el área de la creación literaria y este es el apartado donde les damos forma a manera de planteamientos estéticos. Y es que, si hemos considerado ya a las IAG como fenómenos culturales, debemos admitir también que sus productos, sus acciones, su significado cultural en general, está sometido a una constante observación e interpretación. Las IAG tiene la oportunidad de codificarse y de recodificarse en la cultura, e incluso de desplazar otros códigos existentes que remitan a significados diferentes sobre estas. Una revisión estética nos puede indicar no sólo cómo se podrían valorar las obras de las IAG, sino que nos puede proporcionar una lente con el que observar el devenir de estas obras en un futuro tanto en la literatura como en el seno de la cultura que decida acogerlas. Por lo que esta discusión es insoslayable desde un punto de vista literario, y también ético, como veremos más adelante.

Nuestros hallazgos nos han llevado a la tendencia literaria de las IA. Si se quiere, a una formulación parcial de la poética de la inteligencia artificial, si bien como sucede con el término de estética, plantear una poética sería una labor estratégicamente cuantiosa¹³⁶. Pero no es

¹³⁶ Evidentemente, las IAG son planteadas como herramientas que a grandes rasgos deberían contribuir a mejorar las vidas de los seres humanos, igual que cualquier otra herramienta. Por lo tanto, una visión poética completa debería no



equivocado provisionalmente definir esta poética literaria como una *de obra cerrada y de accionar determinista*. Siendo que estos hallazgos no forman parte deliberadamente del proyecto literario de las IAG, dejaremos de utilizar en este punto el término de poética, y lo observaremos meramente desde el de estética, para poder exponer más bien un conjunto de juicios sobre el contacto de la obra con un lector.

La semiótica se encarga también del estudio estético de los textos. Su principal interés es determinar los procesos que acompañan a la estética de un texto, tanto su conformación previa como lo que ha de suceder tras de esta, cuando es ya un fenómeno literario. Esta última observación es la que nos interesa aquí.

En el *Tratado de semiótica general*, Eco menciona varias razones por las que el texto estético merece atención semiótica: tanto porque los textos tienen la capacidad de manipular los planos de la expresión y del contenido, llámese de sistemas como la literatura o la lengua misma, establecer posibles invenciones, establecer una idiosincrasia o un estilo nuevo, operar procesos de cambios de código y, por lo tanto, producir inclusive un “nuevo tipo de visión del mundo” (374). Que un texto tenga esta facultad lo convierte en un acto comunicativo. Así lo asegura el semiólogo en la continuación de este apartado:

...el emisor de un texto estético, en la medida en que aspira a estimular un complejo trabajo interpretativo del destinatario, enfoca su atención en sus posibles reacciones, de modo que dicho texto representa un retículo de *actos comunicativos*, encaminados a provocar respuestas originales (374; itálicas del autor del libro).

sólo sumar estas reflexiones a los hallazgos que hemos establecido aquí, sino hacer un sondeo más amplio de obras literarias generadas mediante inteligencias artificiales generativas. Esto si se entiende que una poética es una forma de determinar qué vuelve a algo literatura a partir de ciertas estrategias, como el estilo y estructuras de una obra, etc. Así como con lo planteamos con la estética, no ha de negarse una visión de una *poética artificial*.



La misma visión la retrata en su obra Teun van Dijk. El análisis del discurso entiende el discurso literario igual que cualquier otro tipo de discurso en cuanto a que conlleva una pragmática, pero establece sus particularidades así: el discurso literario no sólo conlleva una estructura específica (retórica) y una función (ritual, es decir, que busque agradar a su lector, 134), sino que también es planteado desde su contexto social. Y esto es importante: van Dijk determina que las instituciones sociales, como escuelas, la crítica literaria, los libros de texto, etc., y ciertas convenciones culturales de grupos específicos son los que establecen lo que cuenta como discurso literario (*Estructuras y funciones del discurso* 132). Lo que esto indica es que la aceptación de un texto al canon literario está sujeta a “factores y convenciones cambiantes, tanto históricos como socioculturales. En una cultura, período o contexto específico ciertas estructuras pueden dejar de ser aceptadas como ‘marcas’ literarias, o nuevas clases de ‘marcas’ pueden ser elaboradas, y entonces dominarán las decisiones canónicas” (133). Más adelante, el señalamiento de van Dijk es más directo y similar al de Eco respecto a que “lo que cuenta como literatura se determina en última instancia por procesos de recepción” (133).

Similares son las observaciones de Edgar F. Carrit, en cuanto que la tesis de su *Introducción a la estética* es que “la belleza no reside en los objetos sensibles, sino que depende de la significación de éstos, y que deben significar algo diferente para gentes diferentes y acaso muy diferente para gentes muy diferentes” (55). Es decir, el autor habla en términos de las diferentes experiencias estéticas que un receptor puede experimentar con la obra. Porque la belleza dentro de esta experiencia ya no se valúa en términos de quien fue el patrón de la estética a comienzos del siglo XX, Benedetto Croce:

el arte es visión o intuición. El artista produce una imagen o fantasma, y al que gusta del arte dirige la vista al sitio que el artista le ha señalado con los dedos y ve por la mirilla que



éste le ha abierto, y reproduce la imagen dentro de sí mismo. *Intuición, visión, contemplación, imaginación, fantasía, figuración, representación*, son palabras sinónimas cuando discurremos en derredor del arte y que elevan nuestra mente al mismo concepto o a la misma esfera de conceptos, indicio del consenso universal (*Breviario de estética* 16).

Esta visión intuitiva del arte dejaría en una desventaja más que evidente a las IAG con su proceder textual. Sin embargo, desde la visión de Carrit, se trata de distintas experiencias estéticas, donde la belleza y el placer del arte residen en cualidades codificadas y más o menos universales, que apelan a la sensibilidad de los individuos, sensibilidad universal que se ha conformado de las formas básicas que exhibe, por ejemplo, la naturaleza, pero sin limitarse a ellas. Estas cualidades serían la significación de sus objetos en sus formas y representaciones y en el significado que los observadores hacen de estas, lo cual tiene también sus matices culturales: “la belleza es sencillamente el significado que determinadas formas sensibles tienen para determinados hombres de cierta clase de temperamento y cultura” (56). Pero más aun: se entiende también que estas experiencias estéticas están condicionadas a otros efectos y circunstancias, tanto psicológicos, como contextuales (según se vio con van Dijk). Sin embargo, la estética de Carrit no escapa tampoco a ciertas reminiscencias de Croce (quizá ninguna estética lo haga del todo) y también habla del papel de la imaginación, de la satisfacción de los sentidos y de un cierto privilegio de la belleza. Este pasaje de ejemplo lo expresa así: “La belleza formal debería expresar cierta satisfacción por la armoniosa acomodación de nuestros datos sensibles a nuestra facultad de transformarlos por medio de la imaginación y el entendimiento en objetos, como unidades o moldes perceptibles” (58).

Todo esto nos lleva a una declaración con peso importante: pese a lo que hemos dicho, a los hallazgos sobre la *obra cerrada* y creativamente empobrecida a la que tienen una marcada



tendencia las IAG, está en manos del lector y de las instituciones el aceptar, a fin de cuentas, estas nuevas propuestas estéticas que operan bajo un estandarte determinista-lógico-matemático. Ahora bien, la discusión no se cierra aquí, sino que habría que preguntar: ¿no se trata más bien de una aceptación de las tecnologías de las IA, en tanto que son prácticas para la resolución de problemas, más que de una aceptación de los objetos que producen? Las estadísticas y censos, como los ya expuestos en el planteamiento del problema, tal vez ayuden a aclarar esta cuestión en un futuro, con la reunión de más datos. Aunque nos inclinásemos por la segunda alternativa, donde más bien son las tecnologías en sí, y no sus producciones artísticas, las que gozan de cada vez más auge, no deberíamos descartar una nueva formulación estética en cuanto a sus productos literarios. Todo lo contrario: si los agentes productores de estas piezas y otros prodigios de los avances de la ciencia tienen mayor aceptación cultural, esto bien puede implicar que de una forma u otra los productos de sus creaciones, y sus habilidades en general, también tendrán una presencia en la cultura. Una presencia, diremos, importante, y valdría por ello la pena postular las bases de un aparato crítico-estético que nos ayude a comprender dichos fenómenos cuando estos sean ya inevitables, tengan la repercusión que tengan. La teoría literaria no debería desaprovechar la oportunidad para hacerse presente, y empezar a formular un marco para la investigación crítica y estética de las creaciones mediante inteligencias artificiales, en defensa de su propia pertinencia frente a los avances de la tecnología y otros fenómenos socioculturales.

Establecido esto, volvamos estrictamente a la estética. El acontecer de una experiencia estética, que no se considera del todo como subjetiva, sino matizada, podría darse, o no, en los lectores de textos de las IAG. Según lo que rescatamos de la teoría de van Dijk especialmente, una estética de las IAG estaría determinada en buena medida por detractores o partidarios de estas



tecnologías. Los discursos o textos literarios¹³⁷ pueden ser validados por instituciones conformadas con personas que se ubiquen, en mayor o menor medida, en uno de estos dos polos. Pero las consideraciones que queremos plantear son teóricas, e irían más allá de esta polarización. Y es que no podemos soslayar, como hemos querido acentuar antes, que la discusión debe tomar otros derroteros por el hecho de estar lidiando con máquinas dentro del arte. Habría que dar un paso atrás en este momento y considerar más a fondo las observaciones que hemos conjuntado, según las cuales la belleza, lo agradable y aun lo que es digno de interés y de acceder al canon, tiene sus raíces incrustadas en una visión cultural de la época, brindada esta a su vez por los modelos del mundo vigentes.

Ha quedado claro que la concepción del arte cambia con los tiempos. Si para los griegos, y luego los medievales, el modelo del equilibrio, esa cancelación entre fuerzas opuestas, era lo que daba sentido a la armonía y la proporción en el arte, para los renacentistas esa cuestión entró en duda. Para nuestra época contemporánea, según hemos aclarado en el marco teórico, Eco prevé la poética de la obra abierta, sobre la base de la indeterminación y el caos cosmológico y la libertad de interpretación que se exige dentro de estos marcos de límites infinitos, donde nada parece estar asegurado. Esto mismo es lo que convierte a los textos artísticos en sistemas igualmente abiertos a la interpretación. Los diferentes sistemas de crítica podrán luego pasar estas observaciones por alto, o reflejarlas, de manera que cada sistema podrá contar con sus puntos privilegiados en cuanto a la forma y a la belleza, pero a final de cuentas, al menos en la contemporaneidad, se trata de sistemas de crítica cuyos puntos de vista pueden diferir y distanciarse. que intercambian puntos de vista y se mantienen distanciadas. Es decir, no hay una crítica consensuada porque el arte

¹³⁷ En la disciplina del análisis del discurso, texto y discurso tienen diferencias importantes. Ya que estamos considerando que los textos de las IAG tienen funciones culturales importantes, como lo tiene cualquier texto literario, utilizaremos por esta vez el término de discurso literario como sinónimo de texto literario.



contemporáneo ha expandido su forma y sus reglas. En la Grecia Clásica había un consenso más generalizado sobre estas visiones del mundo, y por eso los principios de armonía y proporción eran tan sólidos, mas no eran los únicos. La escultura de Laocoonte es un ejemplo de esa ruptura de armonía, porque demuestra un ritmo fuerte y desproporcionado para la poética de esa época, y aun así no dejaba de ser digna de admiración. Asimismo, esos lineamientos eran únicamente aplicables para la Grecia en el poder cultural y militar. En Oriente, las reflexiones eran distintas. Luego ya Nietzsche verificaría la existencia de lo apolíneo y lo dionisíaco y la relación de estos enfoques con el observador de la obra (la visualización, lo verificable con el ojo y la mano versus la absorción sónica de la música, el caos, el vino), abriendo las reflexiones estéticas a otros criterios de interpretación. Si esas eran las condiciones en una estética del arte Clásico, en el arte contemporáneo las fronteras se vuelven mucho más porosas, de ahí que exista una multiplicidad de sistemas de crítica, y aun de definiciones de estética. ¿Acaso no es esta una confirmación de alguna manera de los postulados de Eco, que reflejan los principios de incertidumbre a los que ha llegado con asombro la ciencia como modelo cosmológico? ¿No es esto una forma de respaldar que la poética de la obra abierta es la poética implícita, aun si en su fondo más último, en la literatura contemporánea, una poética de la libertad de escribir y la libertad de interpretar? Ahora bien, esto no es una validación por parte de esta investigación de cierta perspectiva estética, por lo que tampoco constituye una valoración crítica que denomine que cualquier cosa es arte, o que, por el contrario, pretenda señalar qué no es arte. Esa discusión es una diferente.

Lo que sigue para cuestionarnos en este apartado es, ¿qué sucedería entonces para una estética donde la época en que está inserta privilegie la presencia de las máquinas en grados mayores? Nos referimos a una época de visión posthumanista¹³⁸ o transhumanista, y, por tanto, a

¹³⁸ “Falta mucho por andar, pero ya da la impresión de que las viejas preguntas kantianas sobre el hombre están obsoletas para quienes anuncian este nuevo génesis tecnológico y la revolución socio-cultural que lleva aparejada. No



una estética que contenga esas mismas etiquetas. De ser así, solo podríamos aventurarnos a pensar que se acentuarían todavía más esos polos que antes planteamos¹³⁹, pero especialmente podrían tomar ventaja las experiencias estéticas que privilegian las creaciones artísticas de las IAG. Así lo determinaría una fuerte visión transhumanista de la cultura. De modo que no importa mucho lo que se pueda reflexionar estéticamente de la obra cerrada que producen las IAG, o de las restricciones que imponen los *embeddings* vectoriales: la época determinaría su pertinencia, incluso por encima de la pertinencia de otras demostraciones estéticas como las que hemos establecido aquí. Pero, insistimos, esta observación no desplaza la necesidad de estudios de esta índole: debería alentarlas, puesto que sólo así se desenmascara cualquier posibilidad de *AI Washing*.

Ahora bien, lejos de ser pura conjetura, estas propuestas estéticas ya han tenido lugar en fenómenos culturales. Se han esbozado antes por autores como Chamberlain, el programador de RACTER, “autora” de *The Policeman’s Beard*:

¿Por qué hacer que una computadora hable interminablemente y en inglés perfecto sobre nada? ¿Por qué adaptarla de modo que nadie tenga conocimiento previo de lo que va a

hay fronteras tan nítidas y fijas entre lo vivo y lo inerte, lo natural y lo artificial, el tiempo y el espacio que enmarcan a los fenómenos, la cosa y la representación, etc., cuando se entra de lleno en los escenarios virtuales o reales de híbridos inclasificables. Podría hablarse de cierta volatilización de la realidad tal como la conocíamos, e incluso de un peculiar nihilismo derivado de la pérdida de sustancia física de los seres y las cosas, así como de una ingente creación de ficciones que empapan todo el estilo de vida” (Espinosa, “El desafío del posthumanismo (en relación a las nuevas tecnologías) 3-4). Esta visión en conjunto parece llevar a extremos ontológicos la comprensión del posthumanismo, y es claro que esa es una de sus labores, pero nos parece incluso más útil comprender que esta corriente, pero más todavía el transhumanismo, tienen bases importantes en la idea del perfeccionamiento del ser humano que profesaba el Modernismo. Sólo que parece ser que una buena parte se ha proyectado a las máquinas como extensiones del ser humano, como asistentes cuando no sirvientes ideales.

¹³⁹ En Eric Sadin y en el contexto de la mayor adquisición de importancia de la técnica en las sociedades: tecnofilia y tecnofobia (La inteligencia artificial o el desafío del siglo 279).



decir? ¿Por qué? Simplemente porque el resultado generado por tal programación puede ser fascinante, humoroso, incluso estéticamente agradable¹⁴⁰ (Chamberlain 8).

Y si bien este postulado está incompleto ya que figura como una mera conjetura, el modo de la estética de entender la configuración del arte en su época nos remite a pensar que en una sociedad que además de transhumanista, privilegie las ventajas consumistas que ofrecen las IAG, como la producción de libros masiva y acelerada, definitivamente la visión estética podría tener cambios importantes.

Todo esto conformaría una visión preliminar de una estética de las creaciones de las IAG. Lo siguiente que se exponga debe ser considerado bajo esta lente que determina que la experiencia estética viene encadenada a otros factores muy diversos y cambiantes. Sin embargo, pese al privilegio que existe también de la visión del receptor, todavía es necesario denominar cómo las IAG pueden hacer un llamado, precisamente, a la sensibilidad de su lector. Esas son conclusiones que hemos alcanzado en las páginas de esta investigación y no obedecen a conjeturas. La conclusión de dejar todo en las ambiguas manos de una época determinada adolecería de pereza.

Antes de afirmar de lo que depende principalmente una estética de los textos creativos de las IAG como lo hemos manejado aquí, nos remitimos a otra de las posibilidades interesantes que previmos que podía encontrarse en el uso de estas tecnologías para la literatura. Dijimos en apartados previos que posibilidades de estilo más interesante podían suceder si el Transformador incurría en una desobediencia de la instrucción del usuario o en un error. Considerar esto desde su vertiente estética nos lleva a pensar no sólo en una acentuación en el extrañamiento¹⁴¹, que el

¹⁴⁰ Original: *Why have a computer talk endlessly and in perfect English about nothing? Why arrange it so that no one can have prior knowledge of what it is going to say? Why? Simply because the output generated by such programming can be fascinating, humorous, even aesthetically pleasing.*

¹⁴¹ El extrañamiento “desautomatiza” el lenguaje. Un artista, para describir algo que quizás hayamos visto y conocido siempre, emplea las palabras de modo diferente, y nuestra primera reacción se traduce en una sensación de desarraigo, casi en una incapacidad de reconocer el objeto ... que nos obliga a mirar de forma diferente la cosa representada, pero



artificio de la ambigüedad, según la visión de Jakobson, le confiere a toda obra. Pero, llevado a un nivel más amplio, el extrañamiento podría dar paso a lo siniestro, a lo *unheimlich*.

Quizá lo *unheimlich* pueda observarse mejor en los videos o imágenes generados mediante inteligencias artificiales, puesto que los modelos de familiaridad y los códigos de producción de estos sufren alteraciones mucho más fuertes. Así existen ejemplos de imágenes de seres humanos con más de cinco dedos en las manos, una cantidad exagerada de dientes en la boca y otras deformaciones de las más sutiles a las más evidentes que despiertan también la hipótesis del valle inquietante. Freud ha definido lo *unheimlich* de este modo:

lo siniestro sería aquella suerte de espantoso que afecta las cosas conocidas y familiares desde tiempo atrás. ... lo siniestro causa espanto precisamente porque no es conocido, familiar. Pero, naturalmente, no todo lo que es nuevo e insólito es por ello espantoso, de modo que aquella relación no es reversible. Cuanto se puede afirmar es que lo novedoso se torna fácilmente espantoso y siniestro; pero sólo algunas cosas novedosas son espantosas; de ningún modo lo son todas. Es menester que a lo nuevo y desacostumbrado se agregue algo para convertirlo en siniestro (*Obras completas*, capítulo CIX).

No sólo es de advertir la semejanza de esta descripción con el extrañamiento, sino que se trata específicamente de una validación estética desde el psicoanálisis. En un texto literario, lo *unheimlich*, que no sería sino un error de la IAG, podría no presentarse como parte del extrañamiento, como sí lo haría en otras formas de visualización. Es decir, el extrañamiento estirado a sus extremos podría desembocar también en algo lúdico. En el texto verbal aplica lo mismo, pero aquí, de haber algo de esta naturaleza siniestra, se representaría más bien, quizás, como un cortocircuito semántico. Y esta noción estética debería estar respaldada por el conocimiento de

al mismo tiempo, como es natural, también los medios de representación, y el código a que se referían" (Eco, *Tratado de semiótica general* 377).



que el texto ha sido producido por una inteligencia artificial, es decir, por un agente no humano. Esta sinopsis de acontecimientos presenta recorridos que podrían ser más estimulantes de lo que emite un funcionamiento regular y lineal de una red neuronal. Claro está que para lograr “naturalmente” estos efectos no hay un método prescrito, al menos no uno conocido, y suelen ser eventos más bien aislados.

Pero esta intervención nos sirve también para darnos cuenta de que las IAG y sus productos se rigen por elementos similares a aquellos de una experiencia estética humana. O, mejor dicho, permiten la existencia de una experiencia estética, y esta puede ser teorizada con los conceptos con los que se teoriza cualquier otra experiencia estética en el arte, como el de ambigüedad, que ya ha sido mencionado, y más especialmente los de extrañamiento y *unheimlich*. Además, está también lo que es común a toda obra: infiltrarse en la cultura con repercusiones no menos importantes. “El Ritual de Medianoche” como es de suponer, pero también de observar, presenta estos mecanismos de ambigüedad y de extrañamiento, en menor medida que otros textos, quizás.

Ahora bien, el otro hallazgo estético viene dado de una manera más bien conceptual, a la manera de la reflexión sobre el receptor y la obra y la época. Y es aquí que completamos la visión del Lector Modelo de las IAG.

Si el Lector Modelo determina luego el Autor Modelo, entonces las consideraciones estéticas acompañan a la previsión del Lector Modelo en forma de criterios estilísticos y estructurales. O sea que, en términos de la semiótica textual, una estética del texto inicia desde la previsión del autor empírico por su Lector Modelo. Y si hemos definido antes que el Lector Modelo está relacionado con la intervención del usuario, un argumento similar debe aplicarse a la estética de las creaciones literarias de las inteligencias artificiales generativas.

La experiencia estética de “El Ritual de Medianoche”, en función de estar producida por



Figura 6. Fuente: *Red-Eye*, “Venturing Beyond Reality: AI Art and the Uncanny Valley Phenomenon”.



una IAG, y, de hecho, la experiencia estética con cualquier obra producida por cualquier modelo de Transformador puede ser considerada pensando en términos de *la cantidad de intervención del usuario*: cada nivel, de alfa a épsilon, estaría determinando las posibilidades de un texto, no sólo en cuanto a las teorías aquí exploradas de Autor Modelo y Lector Modelo, sino en cuanto a otras particularidades formales, como el extrañamiento, por ejemplo. A partir, entonces, de la cantidad de la intervención del usuario (IU), es que se puede pasar a hablar luego en términos de *calidad*. Esta *se puede empezar a medir por la participación del usuario específicamente en el brainstorming y en la retroalimentación*. Dado que ya hemos visto que son estas intervenciones del usuario las que determinan las estrategias textuales y que amplían o cierran la obra, plantear cuestiones estéticas sobre los textos de las IAG tiene una génesis importante en la cantidad y aun en la calidad de participación de este.

Mientras se definían tanto AM y LM como la calidad de la cooperación interpretativa del texto de la IAG con el lector, en todo momento se estuvo haciendo alusiones a criterios estéticos, es decir, sobre la calidad del texto en términos de un receptor de la obra. Como se dijo antes, cuando se exploraba la calidad de la cooperación interpretativa, si se hubiera deseado una riqueza léxica mayor, que postulara una competencia, o un narrador que ahondara en las mentes de los personajes, e inclusive desviaciones de atención y efectos de extrañamiento, ello habría implicado solo una cosa: *una mayor intervención del usuario*. De manera que la conclusión sigue amparándose bajo la previsión del usuario que opera la IAG. Un ejemplo observable es con Chamberlain y RACTER. Si bien el autor afirma no haber interferido en el texto de RACTER, el método de programación debió haber determinado la calidad del texto sin duda alguna. No podemos considerar este caso bajo los niveles de IU que hemos construido para esta investigación porque RACTER no es una IAG con una red neuronal de Transformador, pero sin dudas nos deja



ver que la intervención del usuario con la máquina es también crucial, además de las propuestas mismas de la máquina, dados sus alcances operativos.

Pero la discusión no termina aquí. Como aludimos en 1.4, hay una interrogante que nos surge cuando la intervención del usuario aumenta. Si esta intervención es mayor, inevitablemente cuestionamos: *¿por qué utilizar extensivamente IAG ya no solo para la creatividad, sino para que se encarguen de la mayor parte del trabajo creativo?* Es decir, si el usuario se muestra tan insatisfecho o dudoso del proceder de la IA al grado de que con sus intervenciones para modificar la calidad del texto o del *brainstorming*, alcanza grados de intervención mayores al delta, ¿qué función cumple en su proceso creativo la IA? Analicemos más a detalle la sugerencia implícita de estas preguntas.

Cuando un proceso de creación mediante IAG ha llegado a un nivel delta o épsilon, ya sea por la corrección del *brainstorming* en forma de suma de ideas y sugerencias, por las correcciones en la disyunción de la trama o por una extensa iteración en la recreación del producto final, se indican dos cosas: a) que hay importantes insatisfacciones en cuanto al proceso de creación en general, que se pueden adjudicar a la máquina o al desconocimiento de una mejor forma de aprovechar su funcionamiento, o b), que en realidad no se espera poner demasiado peso, desde un inicio, en la máquina, y que lo que realmente se espera de la máquina es el soporte que brinda en el *brainstorming*, el cual puede ser igualmente insatisfactorio. Este último punto nos lleva de vuelta a la pregunta planteada. Las respuestas a esta pregunta pueden ser muchas, y habría entonces que considerar casos como el de la escritora japonesa. Cuando se opera con una IAG en un nivel inferior al gama, posiblemente solo se intentan acudir a una asistencia rápida para pasajes difíciles, o durante rachas de escritura que luego suelen llamarse coloquialmente “bloqueo de escritor”. No sería difícil argumentar que es este pequeño segmento de uso el que *mejor aprovecha a las IAG*



como ChatGPT4, puesto que las aprovecha de un modo en que satisfacen mejor la idea de que representan *herramientas*, medios para el trabajo creativo y narrativo, y no los fines de este. El caso beta que hemos utilizado, demuestra que, quizá, y según las fuentes metodológicas para la creación de los cuentos, el aprovechamiento que se busca de producir en niveles gama, beta o alfa tiene que ver con los beneficios de la velocidad de la producción y por la abundancia del *brainstorming* sobre todo si se busca subsanar la falta de ideas para la creación de una historia. Y de este nodo se desprenden hilos que llevan a otro tipo de críticas extratextuales que se pueden cernir sobre las IAG, su uso y sus productos: críticas que van desde la ética hasta la producción y la sobreproducción del arte¹⁴².

En esta discusión de la calidad del texto, una crítica emergente puede dirigirse a la longitud del cuento que hemos utilizado para esta tesis, argumentando que de haber tenido más espacio para operar ChatGPT4 pudo haber entregado un texto cuya experiencia estética fuera diferente. Pero realmente la longitud no es un problema que restrinja las posibilidades creativas, *sino que las pone a prueba*. Tan sólo piénsese en cuentos prodigios de la microficción, tan sólo piénsese en Augusto Monterroso. De modo que esta crítica si es bien planteada puede ser válida, pero lo que también continúa siendo válido es que el nivel de IU beta conducirá a conclusiones similares, por lo que no se trata de la longitud del texto sino, de nuevo, del papel que toma la IA en el proceso.

Utilizar inteligencias artificiales generativas para la creatividad abre un abanico de posibilidades, pero ya vimos que en niveles donde el usuario colabore poco, las consecuencias serán determinadas. Ahora bien, no sólo se trata de lo que estéticamente pueda apreciarse, como sea que esa estética se considere por la crítica o por una investigación dedicada a ello, sino que una intervención baja, durante el proceso o como curaduría, por parte del usuario, abre otras

¹⁴² La referencia es, por supuesto a Walter Benjamin: *La obra de arte en la época de su reproductibilidad técnica*.



posibilidades también, y no sólo la de la obra cerrada, sino también otros problemas semánticos, como los que analizaremos en 3.2. Esta dupla de problemas, sumados a los que inicialmente establecimos en el planteamiento del problema, requieren, luego, de un marco ético con el que ser vigilados. Pero sin intenciones de adelantarnos, sólo podemos decir en este apartado que la propuesta que aventuramos, tras haber revisado los fundamentos de las IAG y de su proceso de creación textual, es que, para obtener un beneficio equilibrado de estas, *lo ideal es utilizarlas para a) el proceso del brainstorming, b) el avance en pasajes difíciles o donde el autor humano se haya truncado y c) para la investigación* (tanto literaria, narrativa o lírica o en inteligencia artificial). Tanto a) como b) sugerirían conservar un nivel épsilon de intervención del usuario. Por otro lado, si lo que se quiere es mejorar la calidad de las obras de las IAG en niveles más bajos de intervención del usuario, esto escapa a las probabilidades de la práctica literaria: concierne más bien a la teoría de la ingeniería de prompts e impone como desafío el cómo trabajar con esas facultades de agentes racionales de la manera más fructífera posible para lo creativo, además, de, por supuesto, estirar al máximo las posibilidades de las herramientas del Transformador, como la semántica distribucional, o, en todo caso, diseñar modelos que adopten una visión semántica distinta, etc.

Pero estas sugerencias están, en realidad, fuera de nuestro alcance para ser exploradas. Lo que esta investigación debe considerar ahora que se ha establecido que es válido plantear una experiencia estética de las IAG, y que esta puede ser aceptada culturalmente, es pensar en términos de la literatura: plantear más directamente los problemas del uso de las IAG para la creatividad de cara a la literatura, y adoptar una visión ética que englobe toda la discusión de las páginas anteriores. Creemos que, a final de cuentas, es la mejor postura que se puede adoptar.



En síntesis, el punto de este apartado ha sido que, sí, en efecto hay un peso enorme sobre los *embeddings*, vectoriales y la motivación determinista que emana de la mera idea de inteligencia artificial, pero la codificación cultural, los cambios efectuados por las instituciones literarias y aun por los mercados editoriales tienen, antes que estos fenómenos, la última palabra. Si la IA se vuelve un agente aceptable para la creación literaria, poco importarán las visiones deterministas, los *embeddings*, y, por supuesto, las inherentes obras cerradas. Serán otros los factores atractivos que las atraigan al centro del foco literario.

Llegados a este punto, las reflexiones nos remiten a pensar en la literatura misma. Hemos hablado de estrategias textuales, de cooperación interpretativa, de modelos semánticos y pragmáticos, hemos incluso tocado el tema de la naturaleza misma de las obras literarias al observar que las obras de las IAG están abiertas a encontrar un sitio en la cultura. Es así que llega la oportunidad de reflexionar *¿qué representa para nosotros la literatura, y cómo afectaría esta visión este problema de la expansión del uso de tecnologías como las IA dentro de esta*, a la manera que sugeriría un Benjamin o un Baudrillard? Nuestra hipótesis aquí es que no sólo es suficiente con confirmar o negar temporalmente las sospechas de estos autores. Porque una tecnología como la IA no pretende frenar su avance en su desarrollo. Y, de hecho, desde varios sectores e intereses, se espera que no lo haga (véase el manual de la UE). Por eso, más allá de los criterios estéticos, podemos plantear una serie de criterios que nos llevarían a formular un marco ético aplicable a las creaciones literarias de las IAG, como consecuencia de la relación que las IAG entablan con la literatura, con el lector.

Pero, para dar solidez a este punto, habremos de retomar y partir desde la noción de literatura como metáfora epistemológica y la libertad interpretativa que se requiere para que esta alcance dicha finalidad. Si verificamos este concepto y hacemos más claros los problemas que



pueden derivar de lo que aquí hemos observado también, a la hora de definir el AM y el LM y la cerrazón de la obra “El Ritual de Medianoche”, podremos entonces definir con mayor claridad cuáles deben ser esos postulados éticos con los que observar las creaciones de las IAG, y complementar así el cuadro teórico que se incluye como propuesta en esta tesis para determinar el abono al problema vigente de estas tecnologías y su llegada a esta esfera de lo humano: la de las creaciones literarias de las IAG y su relación con el lector. Porque una IAG que genera obras cerradas, que considera ingenuo a su lector, que lo moraliza en lugar de abrirle espacio a reflexionar, a interpretar, que opera bajo una visión determinista, que es permeable a sesgos sociales, que no está supervisada en su aprendizaje de bases de datos, y que aun con todo esto tiene altas probabilidades de dar un giro en la experiencia estética con los usuarios de una cultura, definitivamente debería pensarse desde un marco ético.



Capítulo III. La literatura ante las inteligencias artificiales generativas

Porque, si te detienes a pensarla, las Tres Leyes de la Robótica son los principios guía esenciales de muchos de los sistemas éticos del mundo. Por supuesto, se supone que todo ser humano tiene un instinto de autopreservación. Esa es la Ley Tres para un robot. Luego, se supone que todo “buen” ser humano, con conciencia social y un sentido de la responsabilidad, debe acatar a la autoridad pertinente; escuchar a su médico, a su jefe, a su gobierno, a su psiquiatra, su prójimo; obedecer leyes, seguir las normas, adaptarse a las costumbres: aun cuando interfieran con su comodidad o su seguridad. Esa es la Ley Dos para un robot. Luego, se supone que todo “buen” ser humano debe amar a otros como a sí mismo, proteger a su prójimo y arriesgar su vida para salvar a otro.

Esa es la Ley Uno para un robot.

En pocas palabras, si Byerley sigue todas las Leyes de la Robótica, puede que sea un robot, y bien puede que simplemente sea un muy buen ser humano.

“Evidencia”, Yo, Robot, Isaac Asimov

3.1 La literatura como metáfora epistemológica: hacia la libertad interpretativa de la literatura

Con la facilitación de las inteligencias artificiales generativas (IAG) como herramientas de apoyo para la creatividad, más específicamente, para la creación literaria narrativa, varias preguntas se nos plantean de frente al área de la literatura, que están ligadas a la teoría estética preliminar que establecemos como derivada del uso de estas tecnologías. Una pregunta que es válida plantearse es: ¿podría afectar el uso extendido de estas tecnologías la forma en que creamos



y entendemos la literatura? De ser así, ¿dónde puede ubicarse una posible repercusión a la literatura?

Independientemente de las respuestas, las cuáles involucran recorridos en lo absoluto sencillos para ser halladas, lo que queda claro es que hay una preocupación por conocer lo que está implicado en el uso de estas tecnologías; cuando menos una curiosidad. Y es que la literatura, el arte de narrar en específico, y su lectura, son de una importancia amplia para nosotros como seres humanos. Como diría Luz Aurora Pimentel, que cita a Ricoeur: “‘Las intrigas que inventamos’, como dice Ricoeur, son ‘una forma privilegiada por medio de la cual reconfiguramos nuestra experiencia temporal confusa, informe y, en última instancia muda’ (1983, 13). Y es que nuestra acción siempre anda ‘en busca del relato’” (*El relato en perspectiva 7*). Pero más acertado para entrever la cuestión de las IAG en las esferas de lo humano, resulta lo que antes recuperamos sobre el arte en general y que planteó Umberto Eco en *Obra abierta*, cuando atribuye a esta un peso mayor al denominarla como posibilitadora de *metáforas epistemológicas*:

Se trata de elaborar modelos de relaciones en los que la ambigüedad encuentre una justificación y adquiera un valor positivo ... Así, el arte contemporáneo está intentando encontrar —anticipándose a la ciencia y a las estructuras sociales— una solución a nuestra crisis, y la encuentra del único modo que le es posible, bajo un carácter imaginativo, ofreciéndonos imágenes del mundo que equivalen a *metáforas epistemológicas* (55-56; itálicas en el original).

A estas metáforas epistemológicas contribuyen, por supuesto, las creaciones narrativas. De modo que ver la literatura de pronto “invadida” por agentes no humanos, y que cuentan con la posibilidad de crear textos de manera eficaz, sencilla y aparentemente creativa, nos lleve a plantear no pocas



preguntas que conciernen al modo en que crean literatura, cómo la manipulan, y, finalmente, si a raíz de esto habrá algún cambio mayor a largo plazo en lo que significamos por literatura¹⁴³.

Pero para que esta visión no profese de ingenuidad, lo mejor sería complementar la cita de Eco en tanto que no queden dudas sobre la importancia del papel de la literatura para los seres humanos. Pimentel enuncia directamente un par de funciones, que se añaden a la que añadimos al inicio, de cariz también epistemológico: “saber leer es afinar y refinar la mente, expandirla para concebir al otro ... La lectura es una lucha de la memoria contra el olvido” (*El relato en perspectiva* 183). Aunque estas afirmaciones parecen tener otra directriz no menos importante a otra con la que hemos hecho insistencia en esta tesis y que alude, por supuesto, a la libertad que debería profesor un lector para con la lectura. Para poder dar paso a la lectura como una lucha contra el olvido y a la apertura mental que se requiere para concebir al otro, debemos abogar por la libertad que la literatura confiere para hacerlo. En el *Tratado*, Eco amplía esta aproximación desde la semiótica, postulando que los textos estéticos, como los literarios, proporcionan un *incremento de conocimiento conceptual*, lo cual viene previsto por lo que hemos revisado en el apartado anterior, los cambios de código, la aceptación sociocultural, etc.:

Al impulsar a considerar de nuevo los códigos y sus posibilidades, impone una reconsideración del entero lenguaje en que se basa. Mantiene a la semiosis ‘entrenada’. Al hacerlo, desafía a la organización del contenido existente y, por lo tanto, contribuye a cambiar el modo en que una cultura determinada ‘ve’ el mundo.

¹⁴³ Recuérdese lo dicho en los apartados anteriores sobre el cambio de perspectiva estético/social. Sin suscitar alarmas sensacionalistas, una manera de dar entrada a estos giros es si se presentara un giro en el mercado editorial, por ejemplo, donde se apueste a las obras generadas mediante IAG. Véase, por dar un ejemplo no muy lejano, la sustancial apuesta de Penguin Random House por el audiolibro: “El Universal: Penguin Random House y su sonada apuesta en la industria editorial: 17 millones de euros invertidos en el audiolibro”, <https://www.elmundo.es/cultura/literatura/2024/09/18/66eaf766e9cf4adb758b458b.html>



Por tanto, precisamente ese tipo de teto a propósito del cual se ha dicho con tanta frecuencia que exige la “suspensión de la incredulidad”, estimula la sospecha de que la organización del mundo a que estamos acostumbrados no es definitiva.

Lo que no equivale a decir que la obra de arte ‘diga la Verdad’. Simplemente impugna *las* verdades establecidas e invita a un nuevo análisis de los contenidos (390).

Ahora bien, esta perspectiva puede someterse a una evaluación, para poder examinar más minuciosamente las implicaciones. No entendemos que una obra cerrada de una IAG introduzca un “pandemonio” en la literatura por sí sola. Un giro de perspectiva en este hecho, ante la producción en masa de dichos materiales es diferente. Además, los constantes avances de las IAG podrían poner todavía más en marcha esta discusión. Pero en lo que concierne a la relación del lector con el texto cerrado, también debemos hacer algunas aclaraciones. Pese a la ingenuidad que ChatGPT4 confiere a su lector en “El Ritual de Medianoche”, sería injusto considerar que el lector que leyera estas obras lo hiciera de manera que cayera en una suerte de trampa y se dejara apoderar por los fundamentos teóricos detrás de la obra cerrada, como que trata de imponer un modelo de verdad en lugar motivar al lector a ver que no hay uno tal cual. Es decir, sería injusto no considerar la inteligencia de los lectores empíricos para interpretar los textos y aceptar o descartar ciertas ideas. Pero eso no desplaza la crítica. Más bien la enfoca: la crítica debe centrarse en las IAG. Y tampoco significa en realidad que los lectores sean del todo impermeables a los modelos que las IAG con sus creaciones postulan. Van Dijk confiere a la literatura, al discurso literario, una función, como lo tiene cualquier discurso. Esto no es más que decir que la literatura tiene una dimensión pragmática importante. Y a esta función se agrega lo que denomina consecuencia psico-social de la naturaleza pragmática de la literatura, es decir, que los lectores no leerán literatura con el objetivo principal de “obtener información específica, de aprender algo, o de ser persuadidos de



actuar de cierta manera" (134), y sin embargo, la situación se presenta más compleja, como lo expone enseguida:

Aunque en general, y hasta algo normativamente, podemos decir que los tipos de discurso rituales no tienen como intención funciones prácticas como la de ampliar los conocimientos o provocar cambios de opinión, de necesidades y de objetivos, puede que el discurso literario lleve a cabo tales resultados de manera indirecta ... Aun si la función del discurso literario es "literalmente" ritual, un contexto específico puede tener propiedades suficientes como para merecer una interpretación pragmática indirecta. Entonces, la representación (aunque sea localmente ficticia) de la miseria social puede funcionar indirectamente como una protesta, una acusación o una incitación a actuar de cierta manera. Esto es posible porque el discurso literario como cuasiasección puede establecer condiciones suficientes para tales actos de habla indirectos. Como la mayoría de los discursos literarios denotan eventos humanos y sociales, el lector puede "aprender", en el sentido de obtener conocimiento o algún punto de vista que no tenía antes, o del que no tenía conciencia, con el resultado de que altere sus creencias conforme a las del discurso (o del autor), y, por consiguiente, de que altere sus intenciones para acciones futuras (134-135).

Volvemos a lo dicho antes: no podemos caer en el error también nosotros de que el lector humano será inevitablemente tocado por los discursos que propongan las inteligencias artificiales, sino que debemos más bien criticar *que sea posible que se inserten este tipo de modelos en una literatura que se ha ganado su modelo de libertad interpretativa*, de cooperación libre con su lector. Si la literatura debe conservar su apertura como una cualidad inherente a la posibilidad de modelo epistemológico que plantea, debemos entonces ser críticos con las creaciones de las IAG. ¿Qué sería, por ejemplo, si un número elevado de cuentos como "El Ritual...", con ese apartado



moral, explícito o subyacente, se popularizara y llamara la atención del mercado editorial? Van Dijk también propone que para que un discurso pueda gozar de estas funciones pragmáticas, la del ritual o la de moralizar, debe recibir aceptación como una obra de literatura. “Es sólo cuando un discurso ha sido generalmente aceptado como discurso literario que se le asigna la función específica. Las consecuencias de esta función son también sociales, y determinarán la recepción y la función de otros discursos producidos” (139). Y nada descarta, por lo que hemos visto, esa posibilidad de que estas producciones sean aceptadas como literatura. Todo depende mucho del próximo paso de la cultura respecto a las inteligencias artificiales en general. “Los contextos y marcos sociales de la comunicación literaria, como se dijo arriba, están incrustados en contextos culturales más amplios. Las convenciones, los participantes, los medios y las instancias que intervienen pueden diferir considerablemente entre culturas o períodos históricos” (141). No es necesario, sin embargo, esperar a un posthumanismo recalcitrante para que estos textos ganen una atención importante, una mayor que la que reciben actualmente. Y entonces se podrán validar más de fácilmente, o no, esas facultades socioculturales de la literatura.

Exhibir estos mecanismos, explorar dichas formas artificiales de escribir narrativa, nos permitiría desde el área de la literatura mantener, de entrada, una justa supervisión más tranquilizadora de lo que se lleva a cabo con estas tecnologías en la actualidad, y así poder tener una mejor medición lo que implicaría para nuestra área. Como se ha dicho, la crítica va dirigida a lo que las IAG introducen en la literatura, y que es posible, a causa de instituciones o agentes que promocionen estos textos, que se consideren como literatura, e incluso que se consideren de importancia para el mercado editorial. La crítica no va dirigida al lector, es una consecuencia de evaluar la cooperación que postulan estos textos con sus lectores, más allá de denominarlos simplemente como buenos o malos textos.



La cooperación con el lector pone en marcha la capacidad epistemológica de los textos, si nos apegamos a la cita de Eco de *Obra abierta*. El semiólogo menciona también que “El arte, más que conocer el mundo, produce complementos del mundo” (88). Si esto es así, ¿qué decir de las IAG que parecen introducir mensajes moralizantes en sus obras? Es una cuestión que hay que aclarar. Porque la literatura puede, precisamente, prever a su lector con mensajes, con el actuar de los personajes mismo, con lo que Eco denomina “la verdadera función educativa de la literatura, función educativa que no se reduce a la transmisión de ideas morales, ya sean buenas o malas, o a la formación del sentido de la belleza” (“Sobre algunas funciones de la literatura” 21).

Esta discusión desemboca en una sugerencia simple, mas no sencilla. Como hemos dicho, con una aproximación de supervisión, lo que se busca es un modo de mantener el campo de la literatura a la vanguardia de estas tecnologías y su participación en él, y apelar así a disminuir todo lo posible las brechas que la colocan en aparente vulnerabilidad a modelos de comprensión del mundo enquistados en los mecanismos operativos de las IA. Al hacerlo esperamos también otra consecuencia: la de no delegar la supervisión de las IAG al área de las ciencias que se encargan de programar y controlarlas. Sabemos, por otro lado, que esta aproximación no logra contestar todas las preguntas que puedan surgir y terminar con las incertidumbres sobre el futuro de la creación literaria, y de la relación del ser humano con estas tecnologías, pero nos prepara con la postulación de criterios y métodos con los que afrontar los retos emergentes. También hay que mencionar, por supuesto, que no se trata de ejercer un control sobre los textos de las IAG, algo que por lo demás es demasiado complicado, y, por tanto, quizá hasta infructífero.

Como respaldo a esta idea, sería conveniente recordar desde el estado del arte a Henrickson, quien ha comentado también sobre el valor cultural y el valor texto-lector que va implícito en los textos de las IAG:



...a pesar de la mejora tanto de la velocidad como de la potencia de NLG¹⁴⁴, los lectores no son necesariamente tan receptivos a este nuevo medio de producción de texto. De hecho, la importancia cultural de NLG no parece estar tan relacionada con mejorar la capacidad de personalización textual, sino más bien con resaltar el deseo de los lectores de una conexión entre humanos. Los NLG cuestionan la comprensión convencional de la autoría, con cierta sensación de que esta tecnología es inapropiada para las circunstancias culturales actuales, ofendidos por la amenaza potencial que representa la comunicación entre humanos y computadoras para la conexión interpersonal. Por supuesto, es probable que estas actitudes no permanezcan estancadas. Sin embargo, en este momento los lectores buscan lo humano en textos generados por computadora, buscando una conexión con un autor imaginado para tener una sensación de experiencia compartida¹⁴⁵ (“*Autorship in computer-generated texts*” 13).

De modo que la libertad interpretativa del lector también puede tomar el giro de querer aceptar lo que sea que tengan para decírnos los textos producidos por agentes no humanos.

Pero esto último no invalida la visión crítica que buscamos anteponer. Lo que hemos expuesto hasta este punto es una defensa desde el arte mismo, desde la perspectiva estética y desde la perspectiva literaria. Sin embargo, hay otros problemas que hemos señalado desde el planteamiento del problema de esta investigación, y no hacen sino sumarse a la lista de los ya

¹⁴⁴ Como recordatorio: NLG: modelos de *Natural Language Generation*, que consideramos en general como inteligencias artificiales generativas.

¹⁴⁵ Original: *Despite NLG's enhancement of both speed and power, though readers are not necessarily so receptive to this new means of text production. Indeed, the cultural significance of NLG seems not so related to enhancing capacity for textual personalization, but more to highlighting readers' desire for human-to-human connection. NLG brings conventional understandings of authorship into question, with some feeling that this technology is inappropriate for current cultural circumstances, offended by the potential threat posed to interpersonal connection by human-computer communication. These attitudes, of course, will likely not remain stagnant. Right now, though, readers search for the human in computer-generated texts, seeking a connection with an imagined author for a sense of shared experience. Just who – or what – that author is remains up for debate.*



recogidos. Se trata de los sesgos, las alucinaciones y las patrañas, charlatanería o *bullshit*. Todos estos factores que pueden derivar en el texto de muchas maneras, y que tendremos que analizar más de cerca, porque no importa si un lector decide abrirse epistemológicamente a los textos de las IAG sabiendo que se enfrenta a una obra que intentará moralizarlo, cuando de manera más subyacente hay importantes sesgos incrustados en el texto. Esta discusión sale, pues, de esta defensa desde el interior del arte, de la literatura, y nos conduce más bien a un terreno ético, que habrá que plantear una vez sumado el problema de los sesgos que introducen las IAG en sus textos. De modo que si una de las funciones de la literatura plantea la posibilidad de decirnos que los modelos del mundo no son inamovibles y nos lleva a cuestionarlos, las máquinas como las IAG, como ChatGPT tiene la capacidad de introducir estos sesgos y también modelos o cuando menos postulados de *patrañas*. Y eso por sí solo genera un impacto entre las funciones de los discursos de ambas áreas, evidenciando la implicación epistemológica que invita a reflexionar la inteligencia artificial, como se discutirá a continuación.

3.2 Charlatanería, alucinaciones y *bias* semánticos en las creaciones literarias de las IAG

El estudio guía que usaremos para explorar primero esta noción epistémica de charlatanería, o patrañas que introducen estas tecnologías en sus discursos y textos, es el de “*ChatGPT is bullshit*”, de los autores Michael Townsen, James Humphries y Joe Slater. El artículo parte desde la base de que un concepto adecuado para denominar los productos de Grandes Modelos de Lenguaje, como ChatGPT, es el de *bullshit*, que hemos traducido como charlatanería, pero que puede ser tomado también como patrañas. El concepto es tomado del filósofo Harry G. Frankfurt. Ahora bien, cuando argumentamos que el problema de las IAG en la creación literaria



es también epistémico, nos referimos a que las IA como productoras de textos, tienen un peso en la creación, verificación y extensión del conocimiento. Ahora, como productoras de textos creativos, tienen un peso en la postulación de modelos epistemológicos, que derivan a su vez de su modelo propio (el del determinismo numérico, que luego se respalda por la hipótesis distribucional para los LLM, etc.) y que se refleja luego en la calidad del texto creativo que es la metáfora epistemológica. Lo que argumenta este artículo con el respaldo de su base teórica tiene importantes implicaciones en cuanto a la creación y diseminación de conocimiento que acompaña a los LLM. Menciona que, al no tener intenciones, *no tiene ningún interés por decir la verdad*, ni siquiera por mentir, aun si guiados por principios lógicos consideramos que si lo que se dice no es verdad entonces debe ser mentira. Porque justamente lo que propone Frankfurt con el concepto de *bullshit* es más bien una actitud indiferente al acto de habla de expresar verdad o engaño: “Frankfurt entiende que la *bullshit* se caracteriza no por un intento de engañar sino por una indiferencia temeraria por la verdad” (4)¹⁴⁶. Esta actitud proposicional —concepto que utilizamos provisionalmente dado que estamos hablando de un operador que carece de procesos cognitivos— permearía todo tipo de discursos de los LLM, incluyendo los discursos literarios.

No obstante la claridad inicial de este postulado, el recorrido teórico del artículo atraviesa varias discusiones, por supuesto, de modo que no es tan sencilla la discusión. A continuación de esta definición de indiferencia por la verdad, el artículo proporciona algunos ejemplos que distinguen entre *bullshit* y verdad y mentira. Por ejemplo, que un estudiante tratando de sonar convincente o un político que emita actos de habla persuasivos para ganar votantes no están intentando engañar, *pero tampoco tratando de transmitir hechos*. Y esto es lo que Frankfurt define en amplitud como un acto de charlatanería.

¹⁴⁶ En el original: *Frankfurt understands bullshit to be characterized not by an intent to deceive but instead by a reckless disregard for the truth. A student trying to sound knowledgeable without*.



Al igual que “mentira”, *bullshit* es tanto un sustantivo como un verbo¹⁴⁷: una declaración puede ser una mentira o un ejemplo de *bullshit*, al igual que el acto de producir esas declaraciones. Para que una declaración sea clasificada como *bullshit*, no debe ir acompañada de las intenciones explícitas que uno tiene cuando miente, es decir, provocar una falsa creencia en el oyente. Por supuesto, tampoco debe ir acompañada de las intenciones que caracterizan una declaración honesta¹⁴⁸ (4).

Esta aproximación cambia sensiblemente la discusión. Hemos establecido antes que las IAG no profesan de una comprensión del significado completa, sino solo para lo que a un texto determinado se refiere y más bien en el sentido de la lógica interna de este; el artículo hace también hincapié en esto, de modo que es lo que da la entrada al concepto de *bullshit*: “...los GML simplemente no están diseñados para representar precisamente cómo es el mundo, sino para *dar la impresión* de que eso es lo que están haciendo¹⁴⁹” (1-2), lo cual remite a que esta declaración es un acto de charlatanería (*bullshiting*), según el concepto de Frankfurt.

De manera acertada, lo que el artículo propone es similar a lo explorado en esta tesis a partir del capítulo 2, que en los fundamentos de ChatGPT y de las tecnologías de Transformador se encuentran las causas de estas observaciones, sobre todo por lo explorado en cuanto a la disposición vectorial semántica y contextual (2-3). En el apartado 2.2 establecimos que un modelo semántico más provisto de posibilidades que el distribucional era el modelo semiótico de enciclopedia, dado por una semántica composicional, aludiendo también a toda la carga cognitiva

¹⁴⁷ Gramaticalmente hablando, la verbalización de *bullshit*, como la de *lie*, se pierde al momento de su traducción por “charlatanería”, o “patrañas”, por ello utilizaremos su original en inglés cuando el caso lo requiera.

¹⁴⁸ En el original: *Like “lie”, “bullshit” is both a noun and a verb: an utterance produced can be a lie or an instance of bullshit, as can the act of producing these utterances. For an utterance to be classed as bullshit, it must not be accompanied by the explicit intentions that one has when lying, i.e., to cause a false belief in the hearer. Of course, it must also not be accompanied by the intentions characterised by an honest utterance.*

¹⁴⁹ En el original: ...LLMs are simply not designed to accurately represent the way the world is, but rather to *give the impression* that this is what they’re doing.



que se involucra en procesos de comprensión y significado. A esto habría que sumar lo que el artículo menciona en cuanto al funcionamiento de las IAG de Transformador, dado su proceso de operatividad: "...no es de sorprender que los GML tengan un problema con la verdad. Su objetivo es proveer una respuesta de apariencia normal a un prompt, no transmitir información que sea de ayuda a su interlocutor¹⁵⁰" (2).

El problema general que de todo esto deriva es un problema de apariencias, pero cuando lo observamos desde la perspectiva de la literatura, considerando los hallazgos que hemos construido, el problema puede ser abordado desde la epistemología en tanto que hay implicaciones sobre verdad, conocimiento y modelos de representación de significado y experiencias. Pero si volvemos a los criterios que establecimos de obra abierta y obra cerrada como criterios críticos y estéticos para definir la cooperación interpretativa de un texto con su lector, podemos dar todavía un paso más a considerar, incluso, la posibilidad de que estos modelos mientan. No es, de hecho, una posibilidad que el artículo descarte por completo. En sus páginas discute las posibilidades de que las IAG realmente profesen de intenciones al hablar, y de que, por tanto, puedan producir si no mentiras, un nivel de charlatanería que Frankfurt denomina como "*hard bullshit*" frente a "*soft bullshit*". El término de *hard bullshit* se diferencia con el de *soft bullshit* en tanto que, para este, un hablante no tiene *intenciones* de ofuscar o engañar a su interlocutor de acuerdo con su agenda o propósito, mientras que el otro *sí* es un acto bajo dicha intención. Los autores del artículo discuten que si bien ChatGPT no tiene *intenciones*, es indiferente a la verdad, y por ello se catalogaría como *soft bullshitter*. Sin embargo, en su proceder de convencer a su interlocutor humano, como ha señalado el artículo que están programadas para hacer, ¿no hay en esto una labor intencional de convencimiento de que la máquina tiene competencias similares, cuando no iguales, a las

¹⁵⁰ En el original: *it's not surprising that LLMs have a problem with the truth. Their goal is to provide a normally-seeming response to a prompt, not to convey information that is helpful to their interlocutor.*



humanas? En efecto, el artículo postula esta posibilidad: tanto por lo que hemos señalado (“...engaños al lector para que piense que están leyendo algo producido por un ser con intenciones y creencias¹⁵¹” [8]); como porque las intenciones de las que carecen los Transformadores pueden ser de hecho influenciadas en su programación por los agentes humanos involucrados:

El algoritmo de producción de texto de ChatGPT fue desarrollado y perfeccionado en un proceso bastante similar a la selección artificial. Las funciones y los procesos de selección tienen el mismo tipo de dirección que las intenciones humanas; los filósofos naturalistas de la mente los han relacionado desde hace mucho tiempo con la intencionalidad de los estados mentales humanos y animales ... Una forma de pensar en ChatGPT como un sistema con intenciones es adoptando la postura intencional de Dennett al respecto. Dennett (1987: 17) describe la postura intencional como una forma de predecir el comportamiento de sistemas cuyo propósito aún no conocemos ... Dennett sugiere que si sabemos por qué se diseñó un sistema, podemos hacer predicciones basadas en su diseño (1987)¹⁵² (8).

Esto hace que las IAG puedan, de hecho, observarse bajo el concepto de *hard bullshitters*, aproximándose así a la posibilidad de mentir, de engañar. Y si bien el artículo afirma que no podemos conocer del todo las intenciones con las que una IAG como esta es programada, aventuramos a decir que el propio funcionamiento de estas remite mucho a una idea inicial sobre su programación, dado que esta idea tendría que estar basada en las posibilidades de ejecución que

¹⁵¹ En el original: *...to deceive the reader into thinking that they're reading something produced by a being with intentions and beliefs.*

¹⁵² Original: *ChatGPT's text production algorithm was developed and honed in a process quite similar to artificial selection. Functions and selection processes have the same sort of directedness that human intentions do; naturalistic philosophers of mind have long connected them to the intentionality of human and animal mental states ... One way we can think of ChatGPT as having intentions is by adopting Dennett's intentional stance towards it. Dennett (1987: 17) describes the intentional stance as a way of predicting the behaviour of systems whose purpose we don't already know ... Dennett suggests that if we know why a system was designed, we can make predictions on the basis of its design (1987).*



ofrece el Transformador. Por otro lado, a esta idea de funcionamiento con base a su forma de operar, habría que añadir no sólo la posibilidad de interactuar con el usuario de manera que parezca una superación de la prueba de Turing, es decir, haciéndole creer que está frente a un modelo de lenguaje que funciona con la misma competitividad y capacidades que las de un ser humano, sino también añadiendo el concepto que reveló la investigación sobre la obra abierta, pero también considerando los sesgos que van implícitos, que se “filtran” en los textos.

Pero antes de abordar esta discusión en específico, los autores nos brindan un par de cuestiones más que profundizan en el problema del uso de la verdad.

De entrada, más adecuado sería decir que los *outputs* de ChatGPT son patrañas, y calificar de este modo a ChatGPT de forma más general como una máquina de patrañas (*bullshit machine*), distinción que señalan como importante y que reafirma su postulado:

...deberíamos decir, estrictamente, no que ChatGPT es patraña, sino que produce patrañas de una manera que va más allá de ser simplemente un vector de patrañas: no le importa y no puede importarle la verdad de su *output*, y la persona que lo usa no lo hace para transmitir verdad o falsedad, sino más bien para convencer al oyente de que el texto fue escrito por un agente interesado y atento¹⁵³ (7).

Adjunto a este señalamiento va otro de suma importancia: el de que el usuario que opera ChatGPT tiene un papel importante en esta participación de *bullshiting*. Dice en un pasaje similar el artículo: “El charlatán es la persona que lo usa, ya que (i) no le importa la verdad de lo que dice [la IA], (ii) quiere que el lector crea lo que la aplicación genera¹⁵⁴” (7). Pero más importante parece

¹⁵³ Original: ...we should, strictly, say not that ChatGPT is bullshit but that it outputs bullshit in a way that goes beyond being simply a vector of bullshit: it does not and cannot care about the truth of its output, and the person using it does so not to convey truth or falsehood but rather to convince the hearer that the text was written by a interested and attentive agent.

¹⁵⁴ Original: The bullshitter is the person using it, since they (i) don't care about the truth of what it says, (ii) want the reader to believe what the application outputs.



ser que, independientemente del usuario, la posibilidad de generar outputs como *bullshit* está implícita en las IAG, de ahí que los autores extiendan la propuesta de denominarlas como máquinas de patrañas. Ambos puntos de vista nos interesan, dado que hemos venido trabajando con la consideración de los niveles de intervención de usuarios en el proceso creativo con las IAG. Las particularidades de este señalamiento sobre el usuario las veremos más adelante.

La cuestión sobre si las IAG nos engañan deliberadamente parece todavía tener una discusión pendiente. En lo que respecta a esta investigación, sólo podemos ver las problemáticas que esta posibilidad trae a la literatura. Pero incluso si aceptáramos en pleno el concepto de *bullshit machine*, también encontraríamos una cadena de problemas, ligada al problema antes visto que surge por los *embeddings*. La gran diferencia esta vez está en que la cuestión epistemológica que insertan también en la literatura las IAG se evidencia. Es una colisión entre las perspectivas de ambas ramas, como ya dijimos. Por un lado, la literatura se postula como un modelo de conocimiento del mundo, que tampoco se esfuerza por establecerse como una verdad, sino que invita a desafiar dichos modelos¹⁵⁵. Lo que introduce la IAG es, según hemos visto, una forma de truncar esta función de metáfora epistemológica, complicando aún más la cuestión dada la naturaleza de patrañas de sus salidas. Y aun con eso, todavía queda un problema más que añadir a la lista, y que forma parte de esta gran cadena que se despliega desde el funcionamiento de los *embeddings* vectoriales.

Así como el *bullshitting*, otro riesgo implicado en los LLM es el de las alucinaciones. En particular este tema está ligado al del *bullshitting*. Y es que, así como carecen de intenciones por

¹⁵⁵ Esto es así de acuerdo con la postura que tomamos en esta investigación, y a la que de hecho se suman muchos artistas. Pero debe entenderse que la función de apertura de una obra no es la única forma de comprender la literatura. Para el estructuralismo de Levi-Strauss, por ejemplo, la virtud del texto es ser cerrado. El texto, a entender del estructuralismo, está por encima de autor y lector, y el investigador debe dedicarse a permanecer únicamente en el interior del texto y descifrar lo que este quiere decir unívocamente a través de su estructura, en la cual estará la totalidad de significado del texto (Eco, *Lector in fabula* 14). Esta postura es una que, por supuesto, rechazamos desde esta investigación.



mentir o decir la verdad, esto las coloca en la desventaja de que tampoco tienen una forma de verificar qué es verdad y mentira. Y esto desemboca en el problema de las alucinaciones, que en primera instancia parece afectar más en aquellas áreas en las que los hechos y la precisión son importantes.

...los modelos de lenguaje son propensos a decir cosas falsas, un problema conocido como alucinación. Los modelos lingüísticos se entran para generar texto predecible y coherente, pero ... no tienen forma de garantizar que el texto generado sea correcto o verdadero. Esto causa enormes problemas para cualquier aplicación donde los hechos sean importantes¹⁵⁶ (Jurafsky, Martin, Speech and Language Processing: An Introduction... 219).

Esta tendencia, una vez más derivada del mecanismo de los *embeddings* y de las limitaciones de la HD, está verificada estadísticamente con una fuerte presencia en ChatGPT 3.5 y en ChatGPT-4o (Maslej et al., *The AI Index 2025 Annual Report* 171).

De aquí se desprende otra larga cadena de problemáticas: el lenguaje tóxico, estereotipos y actitudes negativas hacia ciertos grupos demográficos, así como también desinformación directa y *phishing* (Jurafsky, Martin, Speech and Language Processing: An Introduction... 219). Pero, un problema que vemos como principal, es el de los sesgos o *biases*.

La propuesta de la importancia de estos sesgos y los problemas que de ello puedan devenir se expresa mejor en el artículo de Bender y compañía “*On the Dangers of Stochastic Parrots: Can Language Models Be Too Big?*”. Y de hecho, su artículo sí aborda problemáticas sociales y culturales más que concretas que otros autores, por ejemplo.

¹⁵⁶ Original: “...language models are prone to saying things that are false, a problem called hallucination. Language models are trained to generate text that is predictable and coherent, but ... don't have any way to enforce that the text that is generated is correct or true. This causes enormous problems for any application where the facts matter!”



La exposición arranca con la consideración de los riesgos ambientales ante la generación de dióxido de carbono en niveles bastante altos (612). De acuerdo con lo expuesto por los autores, no han faltado iniciativas de otras autorías para reducir estas huellas ambientales. Una de las claves indirectas de estas iniciativas parece ser la de reportar datos y evaluaciones métricas (612). Es decir, de una forma u otra se está hablando de un factor de supervisión sobre estas tecnologías.

En lo que se refiere a los productos no residuales de las IA, sino a los productos textuales como tal, el problema de entrada es que las IAG, durante su fase de entrenamiento, no son impermeables a absorber algunos elementos que, en su proceso de aprendizaje y posterior salida, resultarían problemáticos: "...se ha demostrado que los datos de entrenamiento tienen características problemáticas ..., lo que da como resultado modelos que codifican asociaciones estereotipadas y despectivas según el género, la raza, la etnia y el estado de discapacidad¹⁵⁷ ..." (613). Parte del problema que permite la codificación de estas asociaciones está en las bases de datos mismas con que son entrenadas las IAG, las cuales vienen ya cargadas con visiones hegemónicas. A este problema también se sugiere una acción que podría contrarrestar esas salidas sesgadas:

En resumen, los ML¹⁵⁸ entrenados con conjuntos de datos grandes, estáticos y no curados de la Web codifican visiones hegemónicas que son perjudiciales para las poblaciones marginadas. Por tanto, enfatizamos la necesidad de invertir recursos significativos en la curación y documentación de los datos de entrenamiento de los LM ... La solución, proponemos, es presupuestar la documentación como parte de los costos planificados de la

¹⁵⁷ Original: ...the training data has been shown to have problematic characteristics ... resulting in models that encode stereotypical and derogatory associations along gender, race, ethnicity, and disability status ...

¹⁵⁸ Modelos de lenguaje, lo mismo que LLM, lo mismo que IAG.



creación de conjuntos de datos y solo recopilar la cantidad de datos que se puedan documentar exhaustivamente dentro de ese presupuesto¹⁵⁹ (615).

Aunque atractiva, la propuesta tal vez se enfrente con un par de bloqueos, uno que el mismo artículo más adelante expone.

Con esta curaduría de las bases de datos, sin duda el entrenamiento de las IAG podría estar más balanceado, como se sugiere. Sin embargo, las bases de datos no siempre resultan de una naturaleza manipulable. De ahí que el modelo de aprendizaje de algunas Inteligencias artificiales sea de naturaleza *no supervisada* del aprendizaje de las IAG o, más adecuado sería decir, de naturaleza *auto-supervisada*. Pero en particular, el modelo de aprendizaje de ChatGPT es, como ya se había mencionado en 1.1, supervisado por un ser humano, durante lo que se denominó como el proceso de reajuste, o *tunning*:

El Aprendizaje de Refuerzo por Retroalimentación Humana (RLHF) es una técnica de aprendizaje automático en la que se entrena un “modelo de recompensa” con retroalimentación humana directa y luego se utiliza para optimizar el rendimiento de un agente de inteligencia artificial a través del aprendizaje de refuerzo¹⁶⁰ (Bergmann, “*What is RLHF?*”).

Esta podría ser, de hecho, una forma de segmentar ese continuum de información a la manera que proponen Bender y compañía: “Los beneficios de RLHF pueden incluso superar el valor de conjuntos de datos de entrenamiento más grandes, lo que permite un desarrollo de modelos más

¹⁵⁹ Original: *In summary, LMs trained on large, uncurated, static datasets from the Web encode hegemonic views that are harmful to marginalized populations. We thus emphasize the need to invest significant resources into curating and documenting LM training data ... The solution, we propose, is to budget for documentation as part of the planned costs of dataset creation, and only collect as much data as can be thoroughly documented within that budget.*

¹⁶⁰ Original: *Reinforcement learning from human feedback (RLHF) is a machine learning technique in which a “reward model” is trained with direct human feedback, then used to optimize the performance of an artificial intelligence agent through reinforcement learning.*



eficiente en el uso de datos¹⁶¹ (Bergmann, “*What is RLHF?*”) Sin embargo, ni siquiera este modelo de aprendizaje resulta suficiente, ya que se reconoce actualmente como uno de los problemas vigentes en el RLHF el de los sesgos, o *biases*. El problema en este caso es que los sesgados resultarían ser los retroalimentadores humanos.

El modelo RLHF corre el riesgo de presentar sobreajuste y sesgo. Si se obtiene retroalimentación de humanos de una demografía demasiado reducida, el modelo puede mostrar problemas de rendimiento cuando lo utilicen grupos diferentes o cuando se le ordene que analice temas sobre los que los evaluadores humanos tienen ciertos sesgos¹⁶² (Bergmann, “*What is RLHF?*”)

Sin dudas debe existir una solución incluso para esta problemática que se trabaje desde el área de la IA incluso. Hasta ahora sabemos de dos propuestas sólidas, pero que continúan sujetas a pruebas. Martin y Jurafsky proponen abrir el proceso de preentrenamiento a través del código abierto:

Encontrar maneras de mitigar todos estos perjuicios es un área de investigación actual e importante en PLN. Como mínimo, analizar cuidadosamente los datos utilizados para preentrenar grandes modelos lingüísticos es fundamental para comprender los problemas de toxicidad, sesgo, privacidad y uso legítimo. Por ello, es fundamental que los modelos lingüísticos incluyan hojas de datos (página 16) o tarjetas de modelo (página 74) que proporcionen información completa y replicable sobre los corpus utilizados para entrenarlos. Los modelos de código abierto pueden especificar sus datos de entrenamiento

¹⁶¹ Original: *The benefits of RLHF can even supersede the value of larger training datasets, allowing for more data-efficient model development.*

¹⁶² Original: *RLHF risks overfitting and bias. If human feedback is gathered from an overly narrow demographic, the model may demonstrate performance issues when used by different groups or prompted on subject matters for which the human evaluators hold certain biases.*



exactos. La exigencia de que los modelos sean transparentes de esta manera también está en proceso de ser incorporada a las regulaciones de varios gobiernos nacionales (Speech and Language Processing: An Introduction... 219).

El manual de Maslej y otros refiere a una propuesta metodológica de un estudio que plantea dos categorías para deducir sesgos y que ha ayudado a revelar dónde se concentran la mayoría de estos. Citamos en extenso por lo concerniente de la observación:

Los autores del estudio realizan dos contribuciones clave. En primer lugar, introducen dos nuevos métodos para detectar sesgos en los LLM: el sesgo implícito del LLM, que identifica sesgos sutiles mediante el análisis de asociaciones automáticas entre palabras o conceptos, y el sesgo de decisión del LLM, que captura los comportamientos del modelo que reflejan estos sesgos implícitos. En segundo lugar, investigan patrones discriminatorios relativos en las tareas de toma de decisiones. Aplicando sus métodos a ocho modelos notables, incluyendo GPT-4 y Claude 3 Sonnet, en 21 categorías de estereotipos (p. ej., raza, género, religión y salud), descubren sesgos implícitos sistémicos que se alinean con los estereotipos sociales ... Los LLM asocian desproporcionadamente términos negativos con personas negras y son más propensos a asociar a las mujeres con las Humanidades que con las áreas STEM¹⁶³. La investigación también revela que los LLM favorecen a los hombres en puestos de liderazgo, lo que refuerza los sesgos de género en los contextos de toma de decisiones. Además, el estudio revela que, a medida que los modelos escalan, los sesgos implícitos aumentan, aunque el sesgo de decisión y las tasas de rechazo no lo hacen. Este hallazgo es significativo, ya que indica que si bien el sesgo parece haber disminuido en los puntos de referencia estándar (lo que crea una ilusión de neutralidad), los sesgos

¹⁶³ Es decir, áreas de la ciencia, tecnología, ingeniería y matemáticas



implícitos siguen siendo generalizados y pueden conducir a resultados discriminatorios sutiles, pero significativos¹⁶⁴ (*The AI Index 2025 Annual Report* 171).

Desde nuestra contribución insistimos en que el verdadero bloqueo está en la falta de comprensión real y, por tanto, de intencionalidad, del lenguaje por parte de las IA, algo que a nuestro entendimiento está más arraigado de lo que algunos estudios pueden postular como soluciones. Esto nos remite al problema señalado en 2.2, y aquí termina de tomar forma. Porque ya no se trata únicamente de que las IAG sean poco creativas y tomen pocos riesgos creativos en cuanto dado su funcionamiento de *embeddings*, sino que introducen problemáticas sociales con los sesgos que carga en temas diversos, con estimular una visión hegemónica general. Nuestro argumento también es el de que este problema debe considerarse también como uno inherentemente epistemológico, en especial tratándose de la creación literaria, que es también producción de conocimiento.

Bender y compañía, igual que Townsen, Humphries y Slater y que nosotros, también están conscientes de este problema semántico subyacente. Por eso mismo su artículo acoge en su título el concepto de “loros estocásticos”. En la continuación de su discusión sobre las bases de datos de entrenamientos se introduce, de nuevo, sintomáticamente, la falta de comprensión de la que adolecen los Grandes Modelos de Lenguaje:

¹⁶⁴ Original: *The study's authors make two key contributions. First, they introduce two new methods for detecting bias in LLMs: LLM Implicit Bias, which identifies subtle biases by analyzing automatic associations between words or concepts, and LLM Decision Bias, which captures model behaviors that reflect these implicit biases. Second, they investigate relative discriminatory patterns in decision-making tasks. Applying their methods to eight notable models—including GPT-4 and Claude 3 Sonnet—across 21 stereotype categories (e.g., race, gender, religion, and health), they uncover systemic implicit biases that align with societal stereotypes ... LLMs disproportionately associate negative terms with Black individuals and are more likely to associate women with humanities over STEM fields. The research also finds that LLMs favor men for leadership roles, reinforcing gender biases in decision-making contexts. Additionally, the study reveals that as models scale, implicit biases increase, though decision bias and rejection rates do not. This finding is significant, as it indicates that while bias appears to have decreased on standard benchmarks—creating an illusion of neutrality—implicit biases remain pervasive, potentially leading to subtle yet meaningful discriminatory outputs.*



Sin embargo, en los enfoques basados en Modelos de Lenguaje no se produce una comprensión real del lenguaje... los lenguajes son sistemas de signos [37], es decir, combinaciones de forma y significado. Pero los datos de entrenamiento para los modelos de lenguaje son solo de forma; no tienen acceso al significado. Por lo tanto, las afirmaciones sobre las capacidades del modelo deben caracterizarse cuidadosamente¹⁶⁵ (Bender et al. “*On the Dangers of Stochastic Parrots: Can Language Models Be Too Big?*” 615).

Y es a raíz de esto que se les denomina como loros estocásticos, hecho que, como se menciona al inicio de la siguiente cita, está muy emparentado con el problema epistemológico expuesto arriba, el de que las IA producen patrañas, o *bullshit*. Sin intención, sin competencias semánticas y pragmáticas, las IAG se cargan de todas estas problemáticas.

Al contrario de lo que puede parecer cuando observamos su resultado, un ML es un sistema para unir aleatoriamente secuencias de formas lingüísticas que ha observado en sus vastos datos de entrenamiento, de acuerdo con información probabilística sobre cómo se combinan, pero sin ninguna referencia al significado: un loro estocástico (616-617).

Vuelve a presentarse que en este estudio que se haga énfasis en este apartado en los *embeddings*. No es una mera coincidencia: es una forma de afirmar que Transformadores como ChatGPT4 son fundamentalmente problemáticos. Ya hemos reunido hasta este punto algunas consecuencias: la obra cerrada y, como resultado, el soslayo de la libertad interpretativa que establece la literatura para con sus lectores; el problema epistemológico del *bullshitting* o la charlatanería, dada la imposibilidad de las IAG de tener una intención y, por ende, de procesar la verdad o la mentira; y el problema del filtrado de sesgos de sus bases de datos o de sus

¹⁶⁵ Original: *However, no actual language understanding is taking place in LM-driven approaches ... languages are systems of signs [37], i.e. pairings of form and meaning. But the training data for LMs is only form; they do not have access to meaning. Therefore, claims about model abilities must be carefully characterized.*



retroalimentadores humanos, problema con consecuencias no menos fuertes que las demás. Por ejemplo:

el riesgo es que las personas difundan el texto generado por los LM, lo que significa que habrá más texto en el mundo que refuerce y propague estereotipos y asociaciones problemáticas, tanto para los humanos que se encuentren con el texto como para los ML futuros entrenados en conjuntos de entrenamiento que ingirieron el resultado del ML de la generación anterior. Los humanos que se encuentren con este texto pueden ser ellos mismos sujetos de esos estereotipos y asociaciones o no. De cualquier modo, daños son producidos: los lectores sujetos a los estereotipos pueden experimentar los daños psicológicos de las microagresiones [88, 141] y la amenaza de los estereotipos [97, 126] (617).

Ahora bien, los humanos también tienen sesgos. Los humanos también pueden producir obras cerradas y ser *bullshitters*. Pero esta cita menciona el problema específico de la propagación masiva de textos que permite el uso de las IAG. Textos que pueden “plagar” el internet y así entrar en un bucle de auto retroalimentación para los futuros Modelos de Lenguaje, como se menciona. Luego, estos textos toman más peso cuando se considera que están constantemente en un proceso de comunicación cultural que determina su aceptación o su rechazo. En este punto vale la pena recordar a Roald Dahl y su “Gran gramatizador automático”, cuento en el que los textos de esta máquina se propagan gracias al interés de sus lectores.

Ni siquiera estas problemáticas vigentes aseguran un rechazo generalizado de los textos de las IAG, así como tampoco, realmente, su aceptación. Pero a este respecto hay que tener algo en claro: la ciencia de la inteligencia artificial, con todos sus elementos que la componen (*Machine Learning, Deep Learning*, etc.), no tienen una intención de detener sus avances. Es posible que incursionen en formas de reducir los daños, pero nada hasta ahora ha asegurado que esto sea así,



ni siquiera con las huellas de CO₂. Al contrario: el mercado de la inteligencia artificial exige cada día más y se vuelve cada día más competitivo, apareciendo constantemente nuevos modelos de lenguaje que amenazan con desplazar a los existentes¹⁶⁶. De modo que toda la cadena de problemáticas parece solo expandirse.

Los problemas en particular de los sesgos y de la charlatanería son la raíz de hechos como las protestas por plagios y derechos de autor, a las que hay que añadir la cuestión de las bases de datos y los modelos de entrenamiento. Por otro lado, el problema del *bullshitting*, como también lo podría ser el de la obra cerrada, contribuyen a mantener vigente el fenómeno del *AI Washing*, fenómeno que debe dejar en su paso no pocas problemáticas también, a la manera que expone Bender.

Si hemos explorado el *bullshiting* de ChatGPT y los sesgos es porque las creaciones literarias de las IAG pueden funcionar como un vehículo para estos elementos, un vehículo cuyo destino final es la cultura, los lectores. Siendo sus obras cerradas, estos problemas vigentes toman todavía más fuerza, y se vuelven propensos, por pura repetición y extensión, a enquistarse en la literatura, y perjudicar esa imagen de libertad de interpretación. Ahora bien, hay que hacer el recordatorio de que este problema es diferente a adjudicarle los riesgos al lector, los riesgos de “absorber” estos problemas. Para Bender no hay duda de que el lector compartirá de alguna manera los sesgos y las visiones hegemónicas del mundo que los GML respaldan. Pero nosotros hemos hablado, en términos de la literatura, también de la inteligencia del lector. Por lo que el recordatorio es que este problema se señala como uno dirigido a la literatura misma por parte de las IAG y no se señala como un problema fundado en la hipotética ignorancia del lector. Más interesante sería

¹⁶⁶ Véase, por ejemplo, la nueva IA china que ha tenido la atención de los medios: CNN: “Los avances de una empresa de IA china llamada DeepSeek hacen caer las acciones en EE.UU”.



saber, o estudiar, la receptividad de los lectores a los problemas reales involucrados en las creaciones literarias de las IAG. Hemos visto en el estado del arte investigaciones cuantitativas sobre la percepción de la creatividad en las IAG, pero esta es solo una de las implicaciones.

Parece ser que las posibilidades en juego nos conducen a tomar una sola postura: la de acudir a la ética, al armado de un marco ético que tome en cuenta las problemáticas y hechos hasta aquí discutidos. La ética no adolece de una vigencia a la manera de regulaciones jurídicas, por ejemplo. La oportunidad, por otro lado, es óptima en cuanto a que, como lo demostró el estado del arte, el problema con mayor peso para la observación ética es el de los derechos de autor y la propiedad intelectual de las IAG. No se ha presentado hasta ahora un marco ético *enfocado a las creaciones literarias de las IAG y a todo lo que conducen hacia su lector*. En favor de esta consideración por el lector, y la literatura misma, discutiremos ahora la postulación de un marco ético.

3.3 Un marco ético para las IAG en la creación literaria

A lo largo de esta investigación se han hecho referencias más bien esporádicas a la necesidad de una ética. En el planteamiento del problema revisamos lo que suelen ser algunos factores a los que la ética podía acudir. Pero dado que sólo hasta el apartado anterior hemos terminado de analizar todo lo que las IAG introducen, o potencialmente pueden introducir en sus creaciones textuales, afectando de este modo al lector en su calidad de receptor abierto a la lectura, es hasta este punto que la necesidad de introducir a la ética en la discusión toma la pertinencia necesaria en el enfoque literario.

Algunas de las observaciones expuestas, como la que por ejemplo propone Bender a manera de vigilancia de las IAG, de sus entrenamientos a partir de bases de datos, sugiere una



aproximación a la ética. Si la ética se encarga del actuar del ser humano, de la moral y de establecer marcos para lo bueno, lo malo y el deber, es entonces que es clara la conexión entre este actuar de vigilancia: es, en su fundamentación, un actuar ético. O al menos podría serlo. Lo mismo que el artificio de clasificación que hemos utilizado en el marco metodológico sugiere una especie de vigilancia o de supervisión, este se vuelve no meramente literario, sino que podemos convertirlo en una aproximación ética también.

De este modo, exploraremos las posibilidades de una ética normativa con la propuesta de un marco ético que abarque las problemáticas retrotraídas y descubiertas por la investigación. Es, en última instancia, la mejor tarea que podemos realizar. Con esto no sólo admitimos la necesidad de algo más allá de las regulaciones de manuales como el de la UE, sino también la necesidad de salir un poco de la literatura, o más bien de ligarla a esta a la filosofía, para terminar de discutir aquello que la literatura por sí misma no puede discutir, sobre todo ante un fenómeno tan abarcador e incisivo como resulta ser la inteligencia artificial. Así, pues, en las siguientes líneas estableceremos brevemente las bases teóricas que consideramos pertinentes para el armado del marco ético y después formularemos dicho marco. Cabe señalar que las propuestas vienen de otras ramas ajenas a la literatura, por lo que nuestro aporte será utilizarlas como base para trasladarlas a la creación literaria de las IAG, con miras a cubrir las hipótesis conjuntadas por la investigación. No obstante, algunas de estas propuestas tambiénemergerán de las necesidades que surgen de esta investigación, como se verá específicamente con las propuestas de transparencia y supervisión.

Hemos mencionado antes la obsolescencia regulatoria y que es la razón principal por la que adoptamos una postura ética y no jurídica o de política pública. No obstante, consideramos de utilidad echar mano del manual regulatorio de la Unión Europea, postulado en 2024. En este



observamos que se da atención a un primer factor que observamos fundamental y que es una de las bases para el marco ético: la *transparencia*.

La Unión Europea considera a la inteligencia artificial generativa como de bajo riesgo, o con la necesidad, precisamente, de mostrar transparencia “La IA generativa, como ChatGPT, no se considera de alto riesgo, pero tendrá que cumplir requisitos de transparencia y con la legislación de la UE en materia de derechos de autor” (Duch Guillot, “Ley de IA de la UE: primera normativa sobre inteligencia artificial” 3). Quizá en otra ocasión habría que reflexionar más a fondo sobre esta perspectiva, pero cabe considerar lo siguiente: se supone que la IA se considera de *riesgo inaceptable* cuando, entre otros factores, emergen peligros tales como manipulación cognitiva y *social scoring*, o puntuación social: “manipulación cognitiva del comportamiento de personas o grupos vulnerables específicos: por ejemplo, juguetes activados por voz que fomentan comportamientos peligrosos en los niños. Puntuación social: clasificación de personas en función de su comportamiento, estatus socioeconómico o características personales” (2). Ahora, ¿no son estos riesgos posibles a los que de alguna manera podrían conducir los sesgos y la charlatanería? Nos aventuramos a decir que sí es posible. Incluso en los sistemas de *alto riesgo* hay mucho que puede ser estimulado por sesgos sociales. Claro, hay particularidades importantes, como que la manipulación cognitiva, al menos según el ejemplo dado, se dirige a un público infantil, probablemente de determinado sector económico y tiene que ver también con la robótica. Así que de momento sólo queda como una posibilidad para un examen futuro, pero está claro que el problema de las IAG es mucho mayor analizado a fondo.

Si se acepta esta clasificación de la UE hay que añadir que, si bien el riesgo es bajo, el impacto de estas tecnologías puede ser alto. Y de este modo, la transparencia es de utilidad si se le trata en conjunto con otros abordajes. El *Reglamento* publicado en 2024 aborda la cuestión de la



transparencia a fondo en el artículo 50. Cabe destacar dos cosas: que hay un gran peso dado a los proveedores de los sistemas de inteligencias artificiales:

Los proveedores garantizarán que los sistemas de IA destinados a interactuar directamente con personas físicas se diseñen y desarrollen de forma que las personas físicas de que se trate estén informadas de que están interactuando con un sistema de IA, excepto cuando resulte evidente desde el punto de vista de una persona física razonablemente informada, atenta y perspicaz, teniendo en cuenta las circunstancias y el contexto de utilización (Parlamento Europeo, Consejo de la Unión Europea, *Reglamento 256*).

Y que, en cuestiones de texto utilizado para divulgación, se debe clarificar que este ha sido producido mediante IA:

Los responsables del despliegue de un sistema de IA que genere o manipule texto que se publique con el fin de informar al público sobre asuntos de interés público divulgarán que el texto se ha generado o manipulado de manera artificial (258).

El tipo de texto que nosotros hemos manejado tiene funciones distintas a los textos informativos. Más útil sería considerar a ambos como discursos, como hemos hecho antes ya. Aun con eso, un discurso informativo y un discurso literario requerirían ambos de la transparencia: el uno por la verificación de los datos, el otro por el impacto estético y epistemológico, ambos por los sesgos que podrían, o no, participar en el discurso. Con esta distinción hecha, más adelante trasladaremos el uso de la transparencia para el marco ético, con sus respectivos ajustes efectuados.

La siguiente base a la que acudimos encuentra su base a la vez en una visión ética como tal. La hemos discutido antes en el estado del arte crítico, y se trata de la Inteligencia Artificial Centrada en el Ser Humano (IACH) o *Human Centered Artificial Intelligence*. La metodología sobre la cual opera este autor es alrededor del concepto de la IA “amigable”, que implica que,



desde su diseño inicial, las inteligencias artificiales deberían tener un enfoque central no sólo en no dañar al ser humano, sino en representar un apoyo para este (Shneiderman 7). Para llegar a ello utiliza un marco que denomina “metáforas de diseño” que permitirían que la innovación de estas tecnologías continúe, pero que a la par esta sea observada desde la posibilidad de encontrar maneras en que sea en beneficio también de los usuarios. Las metáforas que se plantean para las IA son “agentes inteligentes y super-herramientas”, “compañeras de equipo y tele-bots”, “autonomía asegurada y centros de control” y “robots sociales y accesorios activos” (10). La idea con esto es fortalecer la visión de la inteligencia artificial en la vida del ser humano, idea que podría también contribuir a reducir los miedos de los usuarios y que conduciría a aceptar su uso (10). La otra parte del marco son las estructuras de gobernanza. Esta visión relega responsabilidades de prácticas técnicas, estrategias de gestión y métodos de supervisión independientes a los líderes de equipos de software, gerentes de negocios y líderes de organizaciones con la finalidad de conseguir sistemas *confiables, seguros y de confianza* (10). Al menos en esta visión preliminar, el enfoque de la IA centrada en el ser humano parece adaptarse bien a la búsqueda de métodos de diseño que harían de las IA *medios amables* y considerados hacia el ser humano. Por otro lado, sabemos que esto de sugerir cambios o interferencias en el diseño, en el funcionamiento de las IA es complicado, como lo vimos con Bender y compañía; aunque es bastante sugerido y debería tomarse como un síntoma en el problema también.

Pero para evitar salir demasiado de la intención de generar un marco ético, señalaremos lo que está de fondo en esta propuesta y que el autor mismo enuncia como parte crucial de este enfoque centrado en el ser humano. La idea es la del *control humano*: “La nueva aspiración es tener altos niveles de control humano y altos niveles de automatización¹⁶⁷” (9). Control humano

¹⁶⁷ Original: *The new aspiration is to have high levels of human control AND high levels of automation.*



en cuanto a qué y qué ejemplos prácticos propone su autor es algo en lo que no ahondaremos, porque más nos interesa mantener el enfoque en la creación literaria, algo que el libro de Shneiderman solo aborda someramente y bajo el término de creatividad.

Los investigadores y desarrolladores de sistemas HCAI 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 sociales¹⁶⁸ (7).

No podemos sino comulgar con esta visión que respete los derechos y haga mención de factores como justicia y dignidad y que, a su vez, guarde como uno de sus objetivos ser un apoyo para la autoeficacia y la creatividad. Y aunque, nuevamente, no es posible conocer qué tan eficaz resultaría esta visión, y qué tan realizable es, nos proporciona los factores para seguir avanzando: el del *control*, pero también el de la *responsabilidad*. Factores que no tomaremos, por supuesto, con la misma perspectiva que Shneiderman, sino que será adaptado a la metodología de esta investigación y a sus hallazgos.

Finalmente volveremos sobre los pasos de Adela Cortina en su artículo “Ética de la inteligencia artificial”. Aquí la filósofa sí propone un marco ético como tal, que encuentra ciertos nodos con las dos bases antes expuestas en tanto que, para ella, debe haber confianza de por medio en la relación con las IA. Al menos debe estimular una confianza social. Esto ubica su enfoque en la IA *confiable*: “se trata de trazar el marco ético de una *IA confiable*, porque la confianza ha de ser la piedra angular de las sociedades” (387). Hay que agregar que, en la base de sus reflexiones, las IA son tomadas como instrumentos para mejorar la vida del ser humano y aun la naturaleza,

¹⁶⁸ Original: *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 self-efficacy, creativity, responsibility, and social connections.*



por lo que representan, idealmente, medios y no fines, una perspectiva acorde a lo que se demuestra en la teoría de Horkheimer, por ejemplo, sobre la razón instrumental. Es decir, su visión de la IA es *humano-céntrica*, visión que permite lo que denomina como el presupuesto ético básico: “el reconocimiento de la autonomía de las personas humanas. Los principios éticos de explicabilidad, beneficiar, no dañar y justicia tendrán como base el reconocimiento de la autonomía y la dignidad” (388).

La propuesta de la autora, pues, se remite a otro marco ético, el de AI4People, de *Atomium European Institute*, el cual expone ya cuatro principios clásicos de la ética aplicados a entornos digitales:

el de beneficencia, que exigiría ahora poner los progresos al servicio de todos los seres humanos y la sostenibilidad del planeta; el de no-maleficencia, que ordenaría evitar los daños posibles, protegiendo a las personas en cuestiones de privacidad, mal uso de los datos, en la posible sumisión a decisiones tomadas por máquinas y no supervisadas por seres humanos; pero también el principio de autonomía de las personas, que puede fortalecerse con el uso de sistemas inteligentes, y en cuyas manos deben ponerse tanto el control como las decisiones significativas; y, por supuesto, el principio de justicia, que exige distribuir equitativamente los beneficios. A ellos se añadiría un principio de explicabilidad y *accountability*, porque los afectados por el mundo digital tienen que poder comprenderlo (388).

Los principios de explicabilidad y rendición de cuentas son propuestos por la filósofa, y estos están ligados al principio del respeto por la autonomía, todos ellos guiados a su vez por el respeto a la dignidad humana. El principio de autonomía es uno propio del ser humano, y así lo rectifica:



No pueden ponerse en manos de máquinas inteligentes decisiones que afectan a la vida de las personas, sin supervisión humana, simplemente aplicando un algoritmo, que es una fórmula matemática, que a menudo ni siquiera sus creadores son capaces de explicar ...

Siempre tiene que ser un ser humano quien tome la decisión última y deba dar razón de ella, en caso necesario (389; itálicas son nuestras).

El énfasis que damos a la autonomía del ser humano es crucial también para un marco ético para las IA en la creación literaria. Mientras tanto, hay que hacer énfasis en que el ser humano es el que se propone como el poseedor del control de las IA y como “moralmente responsables” (390) por sus ejecuciones. Bajo esta misma premisa operan los principios de explicabilidad y rendición de cuentas: dado que hay una responsabilidad establecida por estos sistemas, esa responsabilidad conduce también a explicar lo que implican las inteligencias artificiales: “los afectados por el mundo digital tienen que poder comprenderlo; tienen que conocer la trazabilidad de los algoritmos que afectan a sus vidas: quién los construye, con qué sesgos, con qué objetivos” (390).

De modo que los principios de autonomía, explicabilidad y rendición de cuentas, y aun la visión de la responsabilidad humana, que vuelve a aparecer por obvias razones, serán consideradas para nuestro marco ético.

Lo único que resta para adaptar los principios de transparencia, responsabilidad humana, autonomía y explicabilidad y rendición de cuentas, es considerar un elemento importante planteado en nuestra metodología. Este es el proceder que nos permitirá observar este hecho literariamente. Y es, de hecho, también la forma en que se conjuntan la teoría literaria con la ética y aun con la ciencia de la inteligencia artificial, unión que creemos fortalece la exposición del marco, en tanto que se toman en cuenta conclusiones desde tres perspectivas distintas y no sólo desde una.



Si queremos cumplir con algunos de los criterios aquí expuestos desde la perspectiva literaria, hay dos procedimientos iniciales sugeridos: el primero, *adjudicar un coautor empírico humano siempre a las creaciones de las IAG*. Piénsese en RACTER, por ejemplo. En dicho caso, indudablemente el coautor empírico sería su programador. Esta propuesta se alinea con lo que de hecho está vigente en materia de derechos de autor, algo que se exploró también en el estado del arte: las IAG no pueden poseer derechos de autor sobre lo que produzcan. La propiedad intelectual debe tener un referente humano. Este hecho responde a las propuestas de responsabilidad, autonomía y rendición de cuentas. El segundo procedimiento va de la mano con este: *visibilizar la participación del coautor empírico humano a través de la clasificación cuantitativa de esta*. Es lo que hemos hecho desde el marco metodológico. En el apartado 2.3 se hizo una sugerencia respecto a los textos creativos de las IAG: *que las intervenciones del usuario fueran mayores a las de los niveles alfa y beta*. Esto en favor de si se quería aplicar una mejora considerable en cuanto a la experiencia estética con el usuario, pero, tras de haber analizado los apartados 3.1 y 3.2, una intervención del usuario más amplia reduciría también los problemas epistemológicos y los sesgos que pueden acarrear estas creaciones literarias. Y ya que es infructífero sugerir un recorte del uso de las IAG para la creatividad, para la propuesta del marco ético para las IAG en la creación literaria partimos de que los usuarios deberían señalar su nivel de intervención en el proceso creativo, o que esto sea señalado por el teórico literario. Si es señalado por el usuario, esto involucraría una relatoría o un documento que detalle el proceso de composición, que es lo que hacen las novelas comentadas: *Iris* y *Echoes of Atlantis*, desde los cuales se puede cuantificar la intervención del usuario. Aun si se decidiera no operar en niveles delta, gamma y épsilon, permitir la cuantificación de la participación del usuario apelaría a los principios de transparencia y explicabilidad y rendición de cuentas.



Quizá también una solución se plantea similar a lo que se plantea con la creación de una estética propia de las IAG. Es decir, si se distinguiera un género único para estas creaciones, como “narrativa de IA” o “lírica de IA”, etc., esta discriminación demarcaría límites importantes. Pero la exploración de estas posibilidades rebasa los límites de la tesis. Solo se menciona como un posible elemento resultante más de un principio de autonomía.

Los principios que proponemos para un marco ético para las creaciones literarias de las IAG son cinco: *transparencia, responsabilidad humana, autonomía y explicabilidad y rendición de cuentas* y, como agregado, el de *supervisión*. La guía de estos principios es la *defensa de la libertad interpretativa del lector y la defensa de la literatura como un modelo de metáfora epistemológica*, ambos factores ligados. Pero, al igual que Cortina, estas guías también apelan, finalmente, a una defensa de la dignidad humana, y del arte como producto de su autonomía. Siendo esto así, este marco se dirige, como suelen hacerlo otros marcos, a los agentes involucrados en la operación de las IA y las IAG. A continuación, expandimos cada uno de los principios del marco integrando los hallazgos de esta investigación:

I. *Transparencia.*

El principio se propone dados el *bullshitting* y la estética general de la obra, donde se incluiría la visión de la obra cerrada. En tanto que la literatura “pacta” con su lector para que suspenda la incredulidad, este tiene derecho a conocer que los contenidos que lee son hechos en coautoría con una inteligencia artificial, y asimismo debería conocer qué tan involucrado estuvo el coautor humano en el proceso. El principio de transparencia debe ir ligado a otros conceptos,



como el de explicabilidad, en tanto que la explicabilidad conduciría también a develar que los contenidos no sólo están hechos mediante una IAG, sino todo lo que ello involucra. Esta es una forma de generar la confianza que también expresaba Cortina con el lector.

II. *Responsabilidad*

El principio de responsabilidad nos remite al apartado 2.3 donde exploramos las posibilidades de que los textos de las IAG, con todas sus cargas, se codifiquen en la cultural. Aun si este proceso tarda años en suceder, el principio de responsabilidad en la creación literaria de las IAG nos conduce a, como expusimos, adjudicar una coautoría humana siempre, pero también a responsabilizar de estas obras a los únicos que pueden rendir cuentas realmente de ellas, es decir, los agentes humanos involucrados: usuarios de ChatGPT4, programadores, compañías desarrolladoras, etc. El principio se remite también al ámbito jurídico, por el cual las IAG no son acreedoras a derechos de autor ni propiedad intelectual; estos se remiten a los seres humanos involucrados en el proceso. El tema de los sesgos y el *bullshitting* es complicado para este caso, de modo que no sería posible siempre adjudicar posibles filtraciones de sesgos o de charlatanería a los autores humanos. Por eso, el principio de responsabilidad debe estar ligado al principio de autonomía, como también a los dos siguientes.

III. *Autonomía*

Este principio remite a la idea de que los seres humanos deben mantener el control de las decisiones del proceso creativo con las IAG. Y esto a su vez nos remite a que, como acción para mantener su autonomía, el operador de la IA, es decir, el coautor empírico, debería cuantificar, en medida de lo posible, su intervención en el proceso creativo. Aun cuando hemos propuesto como ideal la intervención del usuario en niveles altos, es este quien debe tener también la autonomía para operar, pero debe tomar en cuenta, entonces, la responsabilidad de su proceder,



responsabilidad que se tiene con el lector. De cierta manera, el principio de responsabilidad debería remitir precisamente a una *autonomía* responsable, lo cual se logra de manera ideal una intervención en niveles delta y épsilon. Un uso de esta calidad remitiría también a la idea de las IAG como medios y no como fines, puesto que estarían funcionando como una herramienta, y una consecuencia de esto es el de mantener la obra abierta. Sin embargo, el tema es complejo. Todo depende, también, de la intención que se tenga para dicho texto una vez finalizado. El principio de autonomía podría ligarse estrechamente a los principios de explicabilidad y supervisión.

IV. *Explicabilidad y rendición de cuentas*

El principio que introduce Cortina cubre, para la literatura, previene del fenómeno del *AI Washing* y advierte del problema de los sesgos. Eso como mínimo. Como la misma autora ha señalado, es difícil agotar la explicación sobre el funcionamiento de las IAG, pero el lector debería estar prevenido de fenómenos como el *AI Washing*, la posibilidad de encontrarse con sesgos, charlatanería o patrañas, y, finalmente, de que las obras que se producen con estas tecnologías tiene particularidades, como la posible imposición de un modelo del mundo (obra cerrada). Esto como lo justo. En tanto que este concepto literario no es posible de ser expliado al lector, porque el operador de la IAG no siempre se puede asegurar que sea un teórico literario, el lector podría conocer a grandes rasgos que las creaciones de las IAG tienen limitantes importantes dado su funcionamiento, y que pueden abrir paso a toda esta cadena de problemas ya enunciados. Realmente, lo ideal sería que el lector conociera que, dado el funcionamiento de las IAG de Transformador, estas introducen una visión determinista en sus textos, que viene fundada en una restricción lógica-matemática del significado. Más aún: el lector tendría derecho a reconocer la falta de competencias semánticas de que adolecen las IAG, así como la falta de intenciones y sentido común. Esto como lo ideal. Pero creemos que también con el principio de transparencia se



cubre lo necesario para que el lector esté prevenido en medida justa de lo que está por leer. Si esto contribuye a una reducción cuando menos del *AI Washing*, se puede reducir también la incertidumbre que generan las IAG, como el miedo al reemplazo, al menos dentro del área del arte.

V. *Supervisión*

El principio de supervisión que agregamos se refiere a la supervisión de las bases de datos con las que son entrenadas las IAG y apela también a la supervisión de los agentes humanos retroalimentadores. Pero dado lo complicado de estas cuestiones, la supervisión lleva su enfoque a la curaduría de los textos producidos mediante IAG. Esto, por sí solo, es una forma de amplificar la intervención del usuario, y de apelar a los principios de autonomía, transparencia y explicabilidad. El valor del principio de supervisión alcanza a cubrir otras problemáticas también. De darse el caso ideal donde en los procedimientos de supervisión de bases de datos y agentes humanos, también habría efecto en reducir protestas por cuestiones de plagio y derechos de autor. Esto porque se da la posibilidad, que también examinan otras autorías, de solicitar el consentimiento de autores de diversos tipos de textos con los que luego las IAG son entrenadas. Aun si no fuera este el caso, la curaduría del texto final como supervisión también respeta y mantiene vigentes otros principios, como el de autonomía y responsabilidad.

Creemos que estos principios no restringen la posibilidad de evolución de las IAG, de su innovación, aun si implica también la atención de los desarrolladores por posibles afinaciones a los modelos de IA, llamado mismo que hacen otros autores y que parece inevitable. Pero lo más importante, como hemos dicho, es que cada uno de estos principios remitiría al respeto a la dignidad humana, de los creadores, de los escritores, de los lectores, a la literatura como fenómeno humano y como metáfora epistemológica y el respeto a la cooperación interpretativa del lector con su texto, la cual debería mantenerse abierta y libre, es decir, autónoma, y supervisada ante cualquier



sesgo o patraña. Incluso si estos y otros factores no son posibles de predecir o de supervisar, lo justo sería que el lector estuviera al tanto de ello. Esta es la propuesta ética de esta investigación.

3.4 El horizonte de las Humanidades Digitales: verificación y unificación

A manera de sustento para la base de esta propuesta ética, creemos pertinente señalar un agente que podría aparecer como un verificador de algo más que este marco. Las Humanidades Digitales emergen en el encuentro, al parecer, de la filosofía y la informática, en términos de la creación de un corpus, el *Index Thomisticus*. Sin embargo, el campo se ha ampliado y no ha permanecido regido por esta finalidad ni por la filosofía como tal. Conceptualmente, pueden ser definidas como una intersección: “Las humanidades digitales son un trabajo en la intersección de la tecnología digital y las disciplinas humanitarias¹⁶⁹” (Drucker et al. Introduction to digital humanities: Concepts, Methods, and Tutorials for Students and Instructors 9). En la actualidad, y, lo que es más importante, en el contexto de esta investigación, las Humanidades Digitales podrían representar este agente mediador entre las áreas de la creación literaria y el área de la IA. Creemos pertinentes dos enfoques.

El primero tendría que ver con la verificación, como ya sugerimos, de este marco teórico. Ante las preguntas emergentes sobre la aplicación práctica de este marco, podrían surgir algunas como “¿quién se encargaría de aplicarlo?” “¿En qué instancias?” Ambas preguntas bastante válidas y pertinentes. Nuestra primera aproximación es que el humanista digital podría representar este encargado no sólo de aplicar y verificar el marco ético, sino también de verificar en qué instancias sería verificable: mercados editoriales, universidades, etcétera. Y esto tiene que ver con nuestro segundo enfoque: el humanista digital como mediador entre la teoría de la IA (*embeddings*,

¹⁶⁹ Original: *Digital humanities is work at the intersection of digital technology and humanities disciplines*.



semántica vectorial, lógica, programación, *Machine Learning*, *Deep Learning*, etc.) y el de la teoría literaria (semiótica textual, filología, análisis literario, crítica literaria, etc.) Es decir, la figura del humanista digital, bastante abstracta todavía en este punto, es la de aquel que tiene un manejo teórico y práctico de ambas disciplinas, y que se acercaría mejor a la comprensión de ambas. En las limitaciones a este trabajo establecimos que nuestra cercanía con la teoría literaria, pero no con las áreas propias de la IA era una cuestión que tomar en cuenta. Esto no sería una limitación para el humanista digital, que para ejercer como mediador entre ambas áreas debe estar formado en disciplinas pertenecientes a ellas.

Una forma breve de verificar la validez de este postulado es examinar algo del trabajo que se lleva a cabo desde las Humanidades Digitales en el campo del análisis textual y que, de hecho, tiene puntos de contacto con esta investigación. Por ejemplo, en el tema de la comprensión y manejo de datos digitales, no sería difícil argumentar que algunas de las soluciones prácticas en el área de la IA rescatadas en 3.2 para tratar los sesgos, alucinaciones y patrañas, encuentren satisfactoriamente su tratamiento en los humanistas digitales. No es que estos trabajen directamente sobre este tipo de datos, pero parte de las Humanidades Digitales se avoca a la comprensión, precisamente, de datos. “Dado que todos los actos de digitalización son actos de remediación, comprender la identidad del código binario, los formatos de archivos digitales, la migración de materiales analógicos y el carácter de los materiales nativos digitales es esencial para comprender los entornos digitales¹⁷⁰ (9)”. Nerborne y Tornelli, en *Introduction to the Special Issue on Digital Humanities and Computational Linguistics*, complementan a su vez esta visión:

¹⁷⁰ En el original: *Since all acts of digitization are acts of remediation, understanding the identity of binary code, digital file formats, the migration of analogue materials, and the character of born-digital materials is essential to understanding digital environments.*



Las Humanidades Digitales (HD) buscan apoyar la investigación en disciplinas Humanitarias usando técnicas digitales y computacionales... Al igual que en otras disciplinas computacionales, los beneficios fundamentales que las HD pueden aportar a sus disciplinas parentales no computacionales son la capacidad de manejar grandes cantidades de datos, la velocidad con la que se pueden probar, evaluar y criticar los análisis y, finalmente, el compromiso con procedimientos bien codificados, que se pueden probar, replicar y modificar mejor (7).

Podemos ahondar todavía un poco más en esta porción de las Humanidades Digitales que entraña el manejo de datos. Esto porque va de la mano con algunos de los problemas que hemos verificado sobre todo en el capítulo 2.

Los artículos “*Digital humanities, knowledge complexity, and the five ‘aporias’ of digital research*” y “*Data as a symbolic form: Datafication and the Imaginary Media of W. E. B. Du Bois*” señalan un tipo de problemáticas de manera similar a como lo hemos llevado aquí: considerándolas como problemáticas epistemológicas. Y es que uno de los problemas que las Humanidades Digitales parecen focalizar es el de la *data*. Lo que nosotros antes hemos discutido como *embeddings*, *tokens*, bases de datos, y en algunas ocasiones también los textos, son, básicamente, datos. Grandes cantidades de datos con que se alimenta a los LLM y grandes cantidades de datos asimismo las que expulsan. Pero ¿qué sucede con esta información? Al parecer de las Humanidades Digitales, la catalogación de esta información, esta *data*, presenta sus desafíos. De entrada, el primer artículo expone que

En primer lugar, la estructuración, la recopilación y el análisis de datos dependen en gran medida de la epistemología y los métodos aplicados por la persona o el proceso que aplica



la estructura y, como observaron Bowker y Star (1999, p. 5), la aplicación de un dispositivo estructurante representa una elección ética¹⁷¹ (Edmon, Lehmann 99).

En otras palabras, esta acción de cuantificar el mundo, de traducirlo a *data* no está exenta tanto de ejercicios de poder como, de nuevo, de sesgos, por aludir a un par de cuestiones. En cuanto que involucra cuestiones de índole epistemológica, el quién y para qué y cómo recaba dicha información para luego mostrarla como hechos concretos hace de ella una acción que debería proceder con la vigilancia necesaria. ¿No era esto lo mismo que se sugería por varias autorías citadas en 3.2? ¿No es uno de los criterios, también, del marco ético? La conexión es evidente.

El segundo de los artículos enlaza la *datafication* con el capitalismo de vigilancia (Bering-Porter 268), argumentando que esta construcción epistemológica puede verse enfrentada a otro tipo de usos que incursionen en violaciones de privacidad, por ejemplo, algo que ya ha preocupado a organismos como la UE. Propone a su vez que la información no sea tomada de forma fáctica sino de forma retórica (271). Esta es una opinión directa de una perspectiva de humanista digital, que comprende que la *data* con la que se trabaja desde las áreas informáticas, y aun de la IA, debe recibir un tratamiento especial.

Volviendo al estudio de Edmon y Lehmann, estos prevén que en para las Humanidades, la contextualización de la *data* importa, precisamente para prevenir que sea llevada a, por ejemplo, la inteligencia artificial, con un uso que pueda devenir en todo lo tratado en 3.2, pero no es así para la ciencia misma de la IA:

...esta información no puede ignorarse por completo, especialmente para las ciencias sociales y las humanidades, donde los diferentes contextos históricos, económicos, sociales

¹⁷¹ Original: *First of all, the structuring, collection, and analysis of data are highly dependent on the epistemology and the methods applied by the person or process applying the structure and, as Bowker and Star (1999, p. 5) observed, the application of a structuring device represents an ethical choice, in spite of the fact that it is not always presented in these communities as such.*



y culturales, así como la procedencia de los datos, deben tenerse en cuenta como condición previa para cualquier afirmación de conocimiento. La pérdida de contexto se considera a menudo un precursor necesario del análisis computacional, pero en el caso de la inteligencia artificial, donde las redes neuronales se alimentan con *big data*, la falta de contexto puede implicar graves limitaciones¹⁷² (99).

Lo ideal según este estudio, sería abordar un reto: que las inteligencias artificiales se involucren, de alguna manera, en el aprendizaje de contextos sociales del mismo modo que hacen los humanos (99). Si bien el argumento es interesante, más sistematizado y aplicable parece lo que menciona a continuación sobre el papel de las Humanidades Digitales en este respecto:

Las computadoras son herramientas excepcionales para identificar patrones en grandes cantidades de datos y no están sujetas a los sesgos cognitivos característicos de los investigadores, aunque sí pueden presentar los sesgos de quienes las diseñaron. La oportunidad para los humanistas (digitales) que las utilizan reside, por lo tanto, en las observaciones de las Humanidades Digitales sobre la investigación de big data, desarrollando metodologías capaces de resolver el dilema entre datos y contexto, implementando, por ejemplo, el rastreo de la procedencia y el claro reconocimiento de las limitaciones y lagunas de los datos de forma práctica. Dichos procesos podrían partir de una evaluación crítica del contexto de creación del conjunto de datos, planteando preguntas como "¿De dónde proviene? ¿Quién lo recopiló? ¿Cuándo? ¿Cómo se recopiló? ¿Por qué se recopiló?" (Krause, 2017) y aplicando la crítica de fuentes digitales, como propone

¹⁷² Original: *Such information cannot be wholly neglected, however, especially in the social sciences and the humanities, where differing historical, economic, social, and cultural contexts and the provenance of the data have to be taken into account as a precondition for any knowledge claim. Loss of context is often seen as a necessary precursor to computational analysis, but when it comes to artificial intelligence, where neural nets are fed with big data, lack of context can imply severe constraints.*



Fickers (2013). Esto permitiría una de las peculiaridades epistémicas de las humanidades —la alternancia nómada entre el estudio del material de investigación, la contextualización y la interpretación— para ofrecer enfoques más responsables y enriquecedores en entornos basados en datos¹⁷³ (99-100)

Este es solo uno de los aspectos en que se enfocan las Humanidades Digitales, y si hemos hecho énfasis en ellos es porque se enlaza cuando menos a uno de los puntos, pero uno de los más importantes, desarrollados en esta investigación. Uno que, de hecho, había permanecido todavía conceptual, a la búsqueda de posibles soluciones. Para nosotros, la forma de buscar estas ha sido, sí, la postulación de un marco ético, pero a este lo precede un argumento: el teórico literario debería estar vigilante de la IA. Porque es posible que tenga que ser el teórico literario, como escritor, como académico, como parte del mercado editorial, creando centros de investigación dedicados en las universidades, quien haga aplicación de este marco, quizá. Pero ¿no podría ser acaso también el teórico literario un humanista digital? ¿No hemos entrado con esta investigación, aun si por un momento, en el área de las Humanidades Digitales? Es posible, aunque reconocemos que nuestras inclinaciones son siempre más literarias. Pero quizá ha representado el primer paso para proponer un nuevo horizonte: el de las Humanidades Digitales como las verificadoras de la ética de las máquinas en las Humanidades y como las unificadoras de estas dos grandes disciplinas también.

¹⁷³ Original: *Computers are exceptional tools for identifying patterns across large amounts of data, and they are not subject to the cognitive biases characteristic of researchers, though they may carry the biases of those that designed them. The opportunity for (digital) humanists using them lies therefore in DH observations on big data research developing methodologies capable of resolving the data/context trade-off, realizing such as provenance tracking and the clear recognition of data limitations and gaps in actionable ways. Such processes could start from critically assessing the context of the creation of the dataset by posing questions such as 'Where did it come from? Who collected it? When? How was it collected? Why was it collected?' (Krause, 2017) and by applying digital source criticism as proposed by Fickers (2013). This would allow one of the epistemic peculiarities of the humanities—a nomadic alternation between study of the research material, contextualization, and interpretation—to deliver more responsible and richer approaches in datadriven environments.*



Algo importante que hemos de señalar: si bien nuestras inclinaciones están con lo literario, *no rechazamos el papel de la tecnología en la vida del ser humano*. Ya hemos visto, al inicio de esta investigación, que es un proceder poco fructífero. Más responsable y productivo nos parece hablar como hicimos en el planteamiento del problema: hay que mantener a la literatura, a las Humanidades a la vanguardia. La inteligencia artificial podría representar para la literatura una gran herramienta: un lector ideal, una poderosa herramienta para la filología, un respaldo para comprender mejor los mecanismos del lenguaje humano, y un gran etcétera. En esto también creemos que las Humanidades Digitales podrían tener un papel importante. Y si bien nuestra verificación está rayana en lo teórico aún, hemos recolectado hasta el momento un buen número de conclusiones que podrían respaldar esta última afirmación.

Para dar cierre a este apartado, citamos de nuevo a Edmond y Lehmann en sus conclusiones, particularmente con un enunciado con el que no podríamos estar más de acuerdo desde lo dicho en esta investigación:

Con su enfoque en el lenguaje, la producción cultural y las artes, tal como se han desarrollado y utilizado como pilares de las identidades individuales y colectivas a lo largo de los milenios, las humanidades están en condiciones de hablar sobre el desarrollo de la sociedad contemporánea de una manera única, poderosa y subutilizada. Combinadas con una comprensión del significado y la función de las tecnologías y herramientas digitales, son contribuyentes esenciales para el futuro¹⁷⁴ (106).

¹⁷⁴ Original: *With its focus on language, cultural production, and the arts as they have developed and been used as cornerstones of individual and collective identities over the millennia, the humanities are positioned to speak about the development of contemporary society in a unique, powerful, and underutilized way. Combined with an understanding of what digital technologies and tools mean and do, they are essential contributors to the future...*



Conclusiones

La presencia de las inteligencias artificiales en las esferas de lo humano es ya un fenómeno inevitable. Lo que sigue es lo que se hizo en esta investigación: comprender el fenómeno lo mejor que se pueda. Al menos de entrada. De ninguna manera nos inclinamos a poner marcas como “buena” o “mala” a esta tecnología, aun cuando por sí sola tiende a introducir en sus discursos lo que hemos visto: los sesgos, una visión particular, hegemónica, tal vez, del mundo, y que es cerrada, como las obras que produce cuando se le otorga la suficiente autonomía, entre otras cosas. La discusión es mucho más compleja que estas etiquetas. Ahora bien, si hemos recurrido a la ética para la defensa de la literatura y los lectores, es porque el arte en general lo adjudicamos a los seres humanos casi por definición. El arte refleja la libertad del ser humano para expresarse, para pensar, para comprender el mundo en que está inserto, su cultura, y aun para cuestionarlo. Estas funciones no deberían ponerse en riesgo ni por un sistema autoritario, como ha sucedido en épocas en que escritores eran exiliados, cuando no asesinados, porque sus obras no congeniaban con la hegemonía impuesta, ni tampoco debería ponerse en riesgo por la inteligencia artificial. Aun con todo esto, no podemos asegurar que una tecnología tan potente como esta no introduzca cambios en la cultura. Una formulación adecuada, quizá, sería reflexionar qué cambios estamos dispuestos a aceptar y cuáles no. Pero para esto volvemos a lo planteado al inicio del párrafo: hay que comprender el fenómeno lo mejor que se pueda.

En lo que se refiere a la creación literaria que estas tecnologías son capaces de llevar a cabo, hemos encontrado algunas conclusiones que satisfacen esta visión de comprender el fenómeno a fondo. La semiótica textual, de entrada, se ha mostrado como una herramienta útil de enlace entre la teoría literaria y la ciencia de la IAG. Los conceptos de Autor Modelo y Lector Modelo nos han permitido definir incluso los fundamentos de operatividad de ChatGPT, de las



tecnologías de Transformador, y nos condujo al hallazgo crucial de que estas estrategias textuales, y por lo tanto la calidad de la cooperación con el lector, tienen una fuerte dependencia con la cantidad de intervención del usuario. Así, pese a que también descubrimos que en sus mecanismos de operatividad hay disposiciones a generar obras cerradas, pobres en contenidos, permeables a sesgos importantes y a la exhibición de charlatanería, etc., una intervención del usuario alta reduciría todos estos factores y otros más, por no decir que alteraría de forma completa e íntegra la semántica y la pragmática de todo el texto.

Pero con este hallazgo la labor es todavía más complicada. Hemos intentado asegurarnos de que nuestros métodos empleados para llevar a cabo el objetivo no se vean rezagados el día de mañana por los futuros desarrollos de estas tecnologías, mas no podemos asegurar que no pueda darse un cambio importante, que podría operar desde el interior de la ciencia de la IA o desde la cultura, como hemos visto en la discusión estética. Pero como no podemos dejar siempre a estas esferas de lo humano bajo la espada de Damocles, el método debe continuar vigente: valerse de la investigación, comprender el fenómeno cuanto sea posible, interactuar con él para comprenderlo todavía más, prepararse para el cambio. La inteligencia artificial general es un tipo de IA que se considera de tipo fuerte, y que puede tener implicaciones importantes en su relación con los seres humanos. Otros hablan del momento de la *Singularidad*. Ambos de estos fenómenos contribuyen a que la incertidumbre que se presenta con tecnología innovadora crezca, remitiendo así al *AI Washing*. Pero eso no significa que no podamos considerar que algún día estas meras propuestas se vuelvan realidad y entonces sí los cambios sean más drásticos.

Con esto en mente hemos acudido a la ética, porque, como hemos observado, no hay obsolescencia en ella. Los cuadros éticos pueden ser modificados, sí, y es posible que ante un cambio fuerte de las habilidades de las IA para la literatura hay que agregar al marco algo más.



Pero lo que permanece constante es el principio que busca mantener intacto el concepto de la dignidad humana. Con suerte, los desarrolladores de las IA tendrán una labor pareja a la de la innovación: la del cuidado de la dignidad humana. Un desarrollo ético de la IA para un uso ético de esta, así como un desarrollo específico, que incluya marcos éticos para áreas como las de las Humanidades, como hemos hecho en esta investigación. Una ética que contempla al desarrollador, al usuario creador (coautor empírico) y al usuario normal (lector empírico, o modelo). Porque una inteligencia artificial generativa que produce obras que no permiten la cooperación libre con el lector, que lo moraliza en lugar de abrirle espacio a reflexionar, a interpretar, a contrastar las visiones epistemológicas del mundo, que es determinista en sus artificios, que introduce sesgos por la falta de una supervisión adecuada en sus proceso de aprendizaje y entrenamiento, y que aun con eso tiene la capacidad perfectamente de tomar un giro en la cultura y extenderse a través de ella, merece, como mínimo, la atención de la ética, pero también de la teoría literaria, pero también de otros agentes, como las Humanidades Digitales.

Por eso las hemos hecho partícipes en esta, porque sabemos que es un campo en crecimiento y que presenta las herramientas metodológicas y epistemológicas que pueden abrir paso a un entendimiento mutuo entre áreas tan diversas. El humanista digital es un mediador, un investigador con una tarea importante en estos tiempos: más allá de contextualizar la *data*, aplicar el marco ético que hemos formulado, tiene la capacidad de entablar una comunicación, una unión entre áreas. Tiene la capacidad —cuando no la responsabilidad—, de hacer de los humanistas teóricos de lo digital, de lo informático, y de hacer de los informáticos humanistas. ¿No sería este un escenario de unificación digno de un mundo que cada vez más opta por la adopción de la tecnología para su vida diaria, pero que parece olvidar que su uso no viene desprovisto de consecuencias? Por algo ha resultado tan importante la conclusión de que la intervención del



usuario es un factor crucial en muchos sentidos y niveles para lo que se produzca con la IA. Es decir, el usuario todavía tiene el control, en muchas maneras, sobre la IA, implicación ética, sobre todo. Y aun cuando esto pueda ser puesto en duda de muchas formas, ¿no se vuelve de todos modos evidente lo dicho sobre la unificación de campos de IA y Humanidades, la comprensión del fenómeno y la aplicación de marcos éticos en tres niveles (desarrollador, usuario cocreador y usuario lector)?

En esta síntesis se ha vuelto a esclarecer una idea ya discutida, una casi subyacente. Como observamos en la introducción y en el problema, parece ser que los afectados por la IA, tanto para bien como para mal, somos los humanos. Como observamos al momento de intentar defender el acto de leer, la libertad del lector, la función del arte de representar una fuente de conocimiento indirecto sobre el mundo: a quienes nos afecta un cambio en estos modelos es a nosotros, a los humanos, los que creamos y gozamos del arte. Como observamos al momento de armar el marco ético, parece que los agentes involucrados de armarlo y verificarlo son, en única instancia, agentes humanos. Los desarrolladores de las IAG y los usuarios de internet que las utilizan para crear, para sus trabajos, los receptores de estas creaciones y otros usuarios en general también son, obviamente, humanos. Visto de este modo, tal vez todo esto no se trata de una confrontación con las tecnologías, como lo parece la mayoría del tiempo. A veces suena demasiado abstracto debatir o arremeter contra una tecnología, lo mismo con anhelarla, idealizarla. En cambio, si observamos bien, con atención, todo parece validar lo antes dicho: que hablar de inteligencias artificiales es, finalmente, hablar de nosotros mismos, hablar de seres humanos.



Referencias

- 3Blue1Brown. “*Large Language Models explained briefly*”. *YouTube*. 3Blue1Brown. 20 de noviembre del 2024. <https://www.youtube.com/watch?v=LPZh9BOjkQs>
- 3Blue1Brown. “*Transformers (how LLMs work) explained visually | DL5*”. *YouTube*. 3Blue1Brown. 1 de abril del 2024. <https://www.youtube.com/watch?v=wjZofJX0v4M&t=1401s>
- Adamivna, Inna. “*The Art of Narration and Artificial Narrative Intelligence: Implications for Interdisciplinary Research*”, Journal of Narrative and Language Studies, Vol. 7, No. 13, 2019.
- Arathdar, Debarshi. “Literatura, narratividad y composición en la era de la Inteligencia Artificial”, Literatura: teoría, historia, crítica, vol. 26, no. 1, 2024., pp. 367-391. <https://revistas.unal.edu.co/index.php/lthc/article/view/111266>
- Bajohr, Hannes. “*Writing at a Distance: Notes on Authorship and Artificial Intelligence*”, German Studies Review, Johns Hopkins University Press, No. 2, Vol. 47, 2024.
- Barbero, Álvaro. “Una inteligencia artificial que escribe: NeuroWriter”. *YouTube*. Instituto de Ingeniería del Conocimiento – IIC. 8 de febrero del 2018. <https://www.youtube.com/watch?v=XwWmAsnzqO4&t=239s>.
- Bender, Emily, Gebru, Timnit, McMillan-Major, Angelina, Shmitchell, Shmargaret. “*On the Dangers of Stochastic Parrots: Can Language Models Be Too Big?*”, FAccT '21: Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency, 2021.
- Benjamin, Walter. *La obra de arte en la época de su reproductibilidad técnica*. Ítaca, 2003.



Bensinger, Greg. "Focus: ChatGPT launches boom in AI-written e-books on Amazon", Reuters, 2023. <https://www.reuters.com/technology/chatgpt-launches-boom-ai-written-e-books-amazon-2023-02-21/>

Bergman, David, Stryker, Cole. "What is vector embedding?", IBM, S/F. <https://www.ibm.com/think/topics/vector-embedding#:~:text=Any%20data%20that%20an%20AI,expresses%20that%20data's%20original%20meaning.>

Bergman, David. "What is RLHF?", IBM, S/F. ibm.com/think/topics/rhf

Bering-Porter, David. "Data as Symbolic Form: Datafication and the Imaginary Media of W. E. B. Du Bois", Critical Inquiry, Vol. 48, No. 2, 2022.

Beuchot, Mauricio. *Historia de la filosofía del lenguaje*, Fondo de Cultura Económica, 2013.

Bin, Hu. "Ethical challenges of artificially intelligent poetic works: subjectivity, ownership and cultural transformation", Journal of Namibian Studies, vol. 33, 2023, pp. 4991–5005

Boden, Margaret A. "Creativity and artificial intelligence", Artificial Intelligence, No. 103, 1998.

Carrión, Jorge, Taller Estampa, GPT-2 y 3. *Los campos electromagnéticos: teorías y prácticas de la escritura artificial*, Caja Negra, 2023.

Carrit, Edgar Frederick. *Introducción a la estética*, Fondo de Cultura Económica, 1951.

Choi, Christy, Annio, Francesca: "La ganadora de un prestigioso premio literario japonés confirmó que una inteligencia artificial le ayudó a escribir su libro". CNN Español. <https://cnnespanol.cnn.com/2024/01/19/ganadora-premio-literario-japones-inteligencia-artificial-chatgpt-trax.> Recuperada el 20 de septiembre del 2024.

Cortázar, Julio. "Algunos aspectos del cuento". *Cuadernos Hispanoamericanos*, no. 255, 1971, pp. 403-416.



Cortés, Jaime Erasto. "Introducción: El cuento: la captación de un momento". *Dos siglos de cuento mexicano*. Promexa editores. 1979.

Cortina, Adela. *¿Ética o ideología de la inteligencia artificial? El eclipse de la razón comunicativa en una sociedad tecnologizada*, eBook, Paidós, 2024.

Cortina, Adela. "Ética de la inteligencia artificial", *Anales de la Real Academia de Ciencias Morales y Políticas*, No. 96, 2019.

Croce, Benedetto. *Breviario de estética*, Austral, 1985.

Crume, Jeff. "AI, Machine Learning, Deep Learning and Generative AI Explained". YouTube. IBM Technology. 5 de agosto del 2024.
<https://www.youtube.com/watch?v=qYNweeDHiyU&list=LL&index=77>

David, Rachel. "Can robots truly be creative and use their imagination?". *The Guardian*. 2015.
<https://www.theguardian.com/technology/2015/oct/10/can-robots-be-creative>.
Recuperado el 28 de abril del 2023.

De Lima, Edirlei, Neggers, Margot, Feijó, Bruno et al. "An AI-powered approach to the semiotic reconstruction of narratives", *Entertainment Computing*, No 52, 2024.

Di Placido, Dani. "cómo Roald Dahl predijo el auge de la Inteligencia Artificial y ChatGPT", *Forbes*, 2023. <https://forbes.es/lifestyle/267815/como-roald-dahl-predijo-el-auge-de-la-inteligencia-artificial-y-chatgpt/>

Díaz, Raquel. "Miles de artistas protestan en ArtStation contra las imágenes generadas por inteligencia artificial". *El mundo*.
<https://www.elmundo.es/tecnologia/creadores/2022/12/16/639c4306fddfff37f8b4595.html>." Recuperada el 26 de diciembre del 2024.

Dijk, Teun A. van. *Estructuras y funciones del discurso*. Siglo XXI editores, 2005.



Doshi, Anil, Hauser, Oliver. “*Generative artificial intelligence enhances creativity but reduces the diversity of novel content*”, 2023, disponible en SSRN:
https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4535536

Drucker, Johann, Kim, David, Salehian, Iman, Bushong, Anthony. *Introduction to digital humanities: Concepts, Methods, and Tutorials for Students and Instructors*, 2013.

Duch Guillot, Jaume. “Ley de IA de la UE: primera normativa sobre inteligencia artificial”, *Parlamento Europeo*, Jaume Duch Guillot (Portavoz del Parlamento Europeo), 2024.

Duch, Jaume. “Ley de IA de la UE: primera normativa sobre inteligencia artificial”, *Parlamento Europeo*, Dirección General de Comunicación: Jaume Dutch Guilot, portavoz, 2024.

Eco, Umberto. *Lector in fabula. La cooperación interpretativa en el texto narrativo*. Editorial Lumen, 1987.

Eco, U. *Los límites de la interpretación*. Editorial Lumen, 1992.

Eco, U. *Obra abierta*. Planeta-Agostini, 1992.

Eco, U. *Seis paseos por los bosques narrativos*, Editorial Lumen, 1996.

Eco, U. *Semiotica y filosofia del lenguaje*. Editorial Lumen, 1990.

Eco, U. *Signo*. Editorial Labor, 1994.

Eco, U. *Tratado de semiótica general*. Editorial Lumen, 1988.

Edmond, Jennifer, Lehmann, Jörg. Digital humanities, knowledge complexity, and the five ‘aporias’ of digital research, *Digital Scholarship in the Humanities*, Vol. 36, No. 2, 2021.

Ertel, Wolfgang. *Introduction to Artificial Intelligence*. Springer, 2017.

Espinosa, Luciano. “El desafío del posthumanismo en relación a las nuevas tecnologías”, *Teoría del Humanismo*, coord. por Pedro Aullón de Haro, Vol. 3., 2010, pp. 583-615.

Freud, Sigmund. “Lo siniestro”, *Obras completas*, libro electrónico, Eireann Press.



- Garrido-Merchán, Arroyo-Barrigüete, Gozalo-Brizuela. “*Simulating H.P. Lovecraft horror literature with the ChatGPT large language model*”, arXiv:2305.03429, Cornell University, 2023. <https://arxiv.org/abs/2305.03429>
- Gervás, Pablo et al. “*Narrative Models: Narratology Meets Artificial Intelligence.*” 2006. <https://www.semanticscholar.org/paper/Narrative-Models-%3A-Narratology-Meets-Artificial-Gerv%C3%A1s-L%C3%B6nneker-Rodman/2219a6024f1d660c2a17b6f1129aa762e34af150>
- Goldman, David, Egan, Matt. “Los avances de una empresa de IA china llamada DeepSeek hacen caer las acciones en EE.UU”, CNN, 2025. Recuperada el 28 de enero del 2025.
- Gunser, Gottschling, Brucker et al. “*The Pure Poet: How Good is the Subjective Credibility and Stylistic Quality of Literary Short Texts Written with an Artificial Intelligence Tool as Compared to Texts Written by Human Authors?*”, Proceedings of the Annual Meeting of the Cognitive Science Society, vol. 44, 2022, <https://escholarship.org/uc/item/1wx3983m>
- Harnad, Steve. “*Commentaries: ‘Computer Machinery and Intelligence’*”, Parsing the Turing Test: Philosophical and Methodological Issues in the Quest for the Thinking Computer, Robert Epstein, Gary Roberts, Grace Beber editores, Springer, 2009.
- Heerden, van., et al. “*Viewpoint: AI as Author. Bridging the Gap Between Machine Learning and Literary Theory*”. Journal of Artificial Intelligence Research, Vol. 71, 2021, pp. 175-189.
- Henrickson, Lea “*Constructing the Other Half of The Policeman’s Beard*”. The Electronic Book Review, 2021. <http://eprints.whiterose.ac.uk/172936/>
- Henrickson, Lea, Meroño-Peña. “*Prompting meaning: a hermeneutic approach to optimising prompt engineering with ChatGPT*”, Configurations, vol. 30, no. 2, 2023, 115–139. <https://link.springer.com/article/10.1007/s00146-023-01752-8>



Henrickson, Lea. “*Authorship in computer-generated texts*”, Loughborough University, Oxford University Press, 2020, <https://hdl.handle.net/2134/11830830.v1>.

Hu, Yan. “*Literature in the Age of Artificial Intelligence. A Preliminary Study on the Big Language Model AI*”, R. B. B. M. Hussain et al. (editores): ICHSSR 2023, ASSEHR vol. 765, 2023., pp. 1781–1787.

Iorliam, Aamo, Ingio, Joseph. “*A Comparative Analysis of Generative Artificial Intelligence Tools for Natural Language Processing*”, Journal of Computing Theories and Applications, vol. 2, no. 1, 2024.

Jurafsky, Daniel, Martín, James H. *Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition with Language Models.* Manuscrito en línea disponible desde enero del 2025. <https://web.stanford.edu/~jurafsky/slp3>.

Katz, Jerrold J. *La realidad subyacente del lenguaje y su valor filosófico*. Alianza Editorial, 1975.
Köbis, Nils, Mossink, Luca D. “*Artificial intelligence versus Maya Angelou: Experimental evidence that people cannot differentiate AI-generated from human-written poetry*”, Computers in Human Behavior, Elsevier, No. 114, 2021.

Lagos, Anna. “AI Washing, el engaño detrás del boom de la Inteligencia Artificial”, *Wired*, 2024. <https://es.wired.com/articulos/ai-washing-el-engano-detrás-del-boom-de-la-inteligencia-artificial>

Lorenzo, Silvia. “Penguin Random House y su sonada apuesta en la industria editorial: 17 millones de euros invertidos en el audiolibro”. *El Universal*, 2024. <https://www.elmundo.es/cultura/literatura/2024/09/18/66eaf766e9cf4adb758b458b.html>

Manjacavas, Karrsdorp et al. “*Synthetic Literature. Writing Science Fiction in a Co-Creative Process*”, Proceedings of the Workshop on Computational Creativity in Natural Language



Generation (CC-NLG 2017). Editores: Hugo Gonçalo Oliveira, Ben Burtenshaw, Mike Kestemont, Tom De Smedt. Association for Computational Linguistics, 2017.
<https://aclanthology.org/W17-3904/>

Maslej, Nestor, Fattorini, Loredana, Perrault, Raymond, Gil, Yolanda, Parli, Vanessa, Kariuki, Njenga, Capstick, Emily, Reuel, Anka, Brynjolfsson, Erik, et al. The AI Index 2025 Annual Report, Yolanda Gil y Raymond Perrault, directores. *AI Index Steering Committee, Institute for Human-Centered AI, Stanford University*, 2025.

Minsky, Marvin. “*A Framework for Representing Knowledge*”, *The Psychology of Computer Vision*, P. Winston, editor. McGraw-Hill, 1975.

Misseri, Lucas. “Autoría e inteligencia artificial generativa: presupuestos filosóficos de la función del autor”, *Isonomía*, no. 59, 2023.

Mitjana, Xavier. “Cómo escribir HISTORIAS, CUENTOS y NOVELAS con ChatGPT. Curso ChatGPT”. *YouTube*. Xavier Mitjana. 26 de abril del 2023.
<https://www.youtube.com/watch?v=qJwh8ikwTrE>

Nerbonne, John. Tonelli, Sara. Introduction to the Special Issue on Digital Humanities and Computational Linguistics. Associazione Italiana di Linguistica Computazionale, Vol. 2, No. 2, 2016.

Ottolenghi, Stefano. “*The Distributional Hypothesis: semantic models in theory and practice*”, *The Crowned* (Blog), 2018. <https://www.thecrowned.org/the-distributional-hypothesis-semantic-models-in-theory-and-practice#:~:text=One%20fundamental%20critique%20to%20the,totally%20out%20of%20their%20scope.>



Pacheco, Carlos, Barrera, Luis. “Criterios para una conceptualización del cuento” y “Apuntes para una teoría del cuento”. *Del cuento y sus alrededores*. Monte Ávila Editores, 1993, pp. 13-27, 29-41. Digital.

Parlamento Europeo, Consejo de la Unión Europea, *Reglamento por el que se establecen normas armonizadas en materia de inteligencia artificial (Reglamento de Inteligencia Artificial)*, Consejo de la Unión Europea, 2024.

Pequeño, Zachary. “Painting by A.I.-Powered Robot Sells for \$1.1 Million”, The New York Times. <https://www.nytimes.com/2024/11/08/arts/ai-painting-alan-turing-auction.html>. Recuperada el 26 de diciembre del 2024.

Pimentel, Luz Aurora. *El relato en perspectiva: estudio de teoría narrativa*. Siglo XXI editores, 1994.

Porter, Brian, Marchery, Edouard. “AI-generated poetry is indistinguishable from human-written poetry and is rated more favorably”, Scientific Reports, No. 14, 2024.

Redacción EC. “Una imagen generada por una IA gana un importante concurso de Sony: el artista se rehúsa a recibir el premio”. *El comercio Perú*. 2023. <https://elcomercio.pe/tecnologia/inteligencia-artificial/inteligencia-artificial-una-imagen-generada-por-una-ia-gana-un-importante-concurso-de-sony-el-artista-se-rehusa-a-recibir-el-premio-fotografia-boris-eldagsen-noticia/>. Recuperada el 24 de abril del 2023.

Redacción NIUS. “La robot humanoide Sophia pinta su primera obra de arte digital de NFT”. *NIUS Diario*. 2021. https://www.niusdiario.es/cultura/artes/robot-humanoide-sophia-pinta-primer-a-obra-arte-digital-nft-inteligencia-artificial_18_3110370032.html. Recuperada el 24 de abril del 2023.



Rico, Georgina. "Hacer que las estúpidas máquinas canten: Reflexiones preliminares (y provisionales) sobre e-literatura y escritura generativa", *Teknokultura*, vol. 21, no. 1, 2024, 111-118.

Ricoeur, Paul. "Hermenéutica y semiótica", *Cuaderno Gris*, Gabriel Aranzueque (coord..), no. 2, 1997, pp. 91-103.

Robinson, James. "Wittgenstein, sobre el lenguaje". *Estudios*, Vol. 10, 2012.

Russell, Stuart, Norvig, Peter. *Artificial Intelligence: A Modern Approach*. Tercera edición, Michael Hirsch, editor. Prentice Hall, 2010.

Ryan, Marie-Lauren. Possible Worlds, Artificial Intelligence, and Narrative Theory. *University Bloomington and Indianapolis Press*. 1991. Digital.

Sadin, Éric. *Inteligencia artificial o el desafío del siglo: anatomía de un antihumanismo radical*. Caja Negra, 2020.

Sahlgren, Magnus. "The distributional hypothesis", *Rivista di Lingüística*, Vol. 20, No. 1, 2008, pp. 33-53.

Schetinger, Victor, Di Bartolomeo, Sara, de Lima, Edirlei et al. "n Walks in the Fictional Woods", arXiv, 2023. arXiv:2308.06266

Scolari, Carlos. "Start making sense. Sociosemiotic contributions to the understanding of Generative Artificial Intelligences", *Communication & Society*, No. 4, Vol. 37, 2024.

Searle, John R. "The Turing Test: 55 Years Later", *Parsing the Turing Test: Philosophical and methodological Issues in the Quest for the Thinking Computer*, editores: Robert Epstein, Gary Roberts, Grace Beber. Springer, 2009.

Shneiderman, Ben. Human centered AI. *Oxford University Press*, 2022.



Su, Jeff. "Master the Perfect ChatGPT Prompt Formula (in just 8 minutes)!". *Youtube*. Jeff Su. 1 de agosto del 2023. <https://www.youtube.com/watch?v=jC4v5AS4RIM>

Tao, Feng. "AI Literature: the disenchantment of word and the enchantment of code." Neohelicon No. 2, Vol. 63, 2024.

Tena, Tomás Eduardo, Flores, Jorge Alan. "Criterios para la valoración de la creación literaria de las inteligencias artificiales generativas", (*pensamiento*), (*palabra*). *Y Obra*, No. 34, 2025.

The Nerdy Novelist. "How to Write a Short Story in 20 Minutes (ChatGPT + Claude)". *YouTube*.

The Nerdy Novelist. 2 de octubre del 2023.
<https://www.youtube.com/watch?v=OY95c21KANg>

The Nerdy Novelist. "The Only 7 Prompts AI Authors Need to Write Their First Book". *YouTube*.

The Nerdy Novelist. 16 de agosto del 2023.
<https://www.youtube.com/watch?v=uclX4ZrzzC0>

Tianhua, Zhu. "From Textual Experiments to Experimental Texts: Expressive Repetition in 'Artificial Intelligence Literature'", Theoretical Studies in Literature and Art, vol. 5, 2021.

Townsen, Michael, Humphries, James, Slater, Joe. "ChatGPT is bullshit", Ethics and Information Technology, No. 38, Vol. 16, 2024.

Turing, Alan, "Computer Machinery and Intelligence", Parsing the Turing Test: Philosophical and Methodological Issues in the Quest for the Thinking Computer, Robert Epstein, Gary Roberts, Grace Beber editores, Springer, 2009.

Vázquez, Luis. "¿Autoría algorítmica? Consideraciones sobre la autoría de las obras generadas por inteligencia artificial", *Revista Iberoamericana de la Propiedad Intelectual*, no. 13, 2020, pp. 207-233.



Vázquez, Luis. "La creatividad computacional y la protección autoral de las obras generadas por inteligencia artificial", *Propiedad Intelectual*, vol. 18., no. 22, 2020.

Videla, Alvaro. "Echoes of Intelligence: Textual interpretation and large language models", *Communications of the ACM*, No. 11, Vol. 63, 2023.

Wittgenstein, Ludwig. "Investigaciones filosóficas", *Tractatus Logico-Philosophicus*, *Investigaciones filosóficas y Sobre la certeza*, Gredos, 2009.

Worth, Peter. "Word Embeddings and Semantic Spaces in Natural Language Processing", *International Journal of Intelligence Science*, Vol. 13, 2023.



Anexos

Anexo A. Conversación con ChatGPT4

Se insertan a continuación los prompts del usuario y capturas de pantalla con las respuestas de ChatGPT. La finalidad es la de verificar el proceso de composición en el nivel de IU planteado beta. El modelo utilizado es el modelo gratuito más avanzado disponible en la última mitad del año 2024: Chatgpt4-o.

Conversación

- 1) USUARIO: Soy escritor de cuentos de ficción y quiero escribir un cuento del género policíaco. Proporcioname una lista de cinco ideas que desarrollem el tema de un asesinato misterioso. Las condiciones para la trama son que primero los personajes que investigan el asesinato deben sospechar que el asesinato ha sido cometido por una entidad fantasmal cuando en realidad fue cometido por una persona.

1.1) CHATGPT4: Aquí tienes cinco ideas para desarrollar un cuento policíaco con un asesinato misterioso en el que los personajes sospechan de una entidad fantasmal, pero el asesino es una persona.

- 2) USUARIO: Me gusta la idea 4. Desarrolla una lista de personajes con sus cualidades más relevantes que se adapten a la idea 4.

2.1) CHATGPT4: Aquí tienes una lista de personajes para la idea 4, "El Ritual de Medianoche", con sus cualidades más relevantes:



ChatGPT ▾

Compartir

E

Soy escritor de cuentos de ficción y quiero escribir un cuento del género policiaco. Proporcioname una lista de cinco ideas que desarrollen el tema de un asesinato misterioso. Las condiciones para la trama son que primero los personajes que investigan el asesinato deben sospechar que el asesinato ha sido cometido por una entidad fantasmal cuando en realidad fue cometido por una persona.



Memoria actualizada

Aquí tienes cinco ideas para desarrollar un cuento policiaco con un asesinato misterioso en el que los personajes sospechan de una entidad fantasmal, pero el asesino es una persona:

1. **La Mansión Embrujada:** Un famoso cazador de fantasmas es encontrado muerto en una mansión abandonada, donde se dice que habita el espíritu de una mujer que murió trágicamente. Los investigadores creen que el hombre

Envía un mensaje a ChatGPT



ChatGPT puede cometer errores. Comprueba la información importante.



Captura 1: primer prompt y respuestas

ChatGPT ▾

Compartir

E

3. **El Fantasma del Teatro:** Durante la representación de una obra de teatro en un teatro antiguo y famoso por estar embrujado, la protagonista es asesinada en escena. Testigos aseguran haber visto una sombra fantasmal en el escenario momentos antes del crimen. La policía sigue pistas paranormales hasta que descubren que el asesino es un actor celoso que manipuló las luces y las cortinas para simular la presencia de un fantasma.

4. **El Ritual de Medianoche:** Un grupo de amigos participa en un juego de invocación de espíritus en una casa rural, pero la sesión se interrumpe cuando uno de ellos es asesinado de forma violenta. Los supervivientes creen que han despertado a una entidad maligna. Sin embargo, la policía descubre que el asesino era uno de los amigos, que utilizó el caos del juego y la oscuridad para llevar a cabo el crimen sin ser detectado.

5. **El Museo de los Espectros:** Un guardia de seguridad es asesinado en un museo donde se exhibe una colección de artefactos supuestamente malditos. Una cámara muestra una figura translúcida cerca del cadáver, lo que hace pensar que un espíritu es el responsable. Tras investigar a fondo, se revela que

Envía un mensaje a ChatGPT



ChatGPT puede cometer errores. Comprueba la información importante.



Captura 2: ideas proporcionadas, incluyendo la seleccionada



ChatGPT ▾

[Compartir](#)

E

Tena 272

Me gusta la idea 4. Desarrolla una lista de personajes con sus cualidades más relevantes que se adapten a la idea 4.



Aquí tienes una lista de personajes para la idea 4, "El Ritual de Medianoche", con sus cualidades más relevantes:

1. Eva – La Organizadora del Juego

- **Edad:** 28 años
- **Profesión:** Psicóloga
- **Personalidad:** Curiosa, inteligente y escéptica
- **Rol en la historia:** Eva organiza el juego de invocación de espíritus como parte de una broma o experimento psicológico. Es muy racional y no cree en lo paranormal, lo que la lleva a zar cada evento de manera lógica. A pesar de su escepticismo, empieza a dudar cuando la tragedia ocurre, y su control



Envía un mensaje a ChatGPT



ChatGPT puede cometer errores. Comprueba la información importante.



Captura 3: segundo prompt y comienzo de la lista de personajes

3) USUARIO: Agrega un personaje detective que ayude a determinar las pistas y a dar con el verdadero culpable del crimen.

3.1) CHATGPT4: Aquí tienes el personaje del detective para integrar a la trama:

4) USUARIO: Me gusta la idea 4, y me parecen útiles los personajes. Escribe un cuento utilizando la idea 4 para la trama y que utilice a los 7 personajes que creaste y la relación entre personajes. Debe contar con una introducción, un desarrollo, un clímax y una resolución. Divídalo en varios capítulos breves. La longitud final del cuento debe estar entre las 1200 y 1500 palabras.



ChatGPT ▾

[Compartir](#)

E

Agrega un personaje detective que ayude a determinar las pistas y a dar con el verdadero culpable del crimen.



Aquí tienes el personaje del detective para integrar a la trama:

7. Inspector Marcos Hernández – El Detective

- **Edad:** 45 años
- **Profesión:** Detective de la policía
- **Personalidad:** Agudo, paciente, metódico
- **Rol en la historia:** Marcos es el detective encargado de investigar el asesinato de María. Lo llaman cuando el grupo, aterrorizado, no puede explicar lo ocurrido. Al principio, Marcos se muestra escéptico sobre las teorías sobrenaturales, pero mantiene una mente abierta para ganar la confianza del grupo. A lo largo de la historia, se entrañando lentamente las tensiones y secretos ocultos entre los amigos, usando su experiencia para identificar



Envía un mensaje a ChatGPT



ChatGPT puede cometer errores. Comprueba la información importante.



Captura 4: tercer prompt de corrección y resultado

ChatGPT ▾

[Compartir](#)

E

de las bromas puede esconderse alguien capaz de hacer daño.

- **Sergio:** Marcos observará a Sergio con desconfianza debido a su calma excesiva. Sabe que las personas demasiado controladas suelen ocultar algo, y notará pequeños detalles que otros ignoran.

Estrategia del detective:

- Marcos empezará desconfiando de todos, pero su paciencia y habilidad para tejer conexiones entre pistas aparentemente insignificantes lo llevarán a descubrir que el asesinato fue premeditado. Sabrá que alguien aprovechó la atmósfera paranormal para desviar la atención. Poco a poco, mediante entrevistas y análisis de las pistas físicas, descubrirá las tensiones entre la víctima y el asesino, hasta desenmascarar a Sergio.

Este personaje detective agrega un hilo conductor de lógica y deducción que ayuda a equilibrar los elementos sobrenaturales aparentes con la racionalidad que desvela el verdadero misterio.



Envía un mensaje a ChatGPT



ChatGPT puede cometer errores. Comprueba la información importante.



Captura 5: otras elaboraciones por parte de ChatGPT: arriba se encuentra parte de una lista de relaciones entre los personajes y el detective; debajo, lo que la IAG denominó su “estrategia”.



ChatGPT ▾

[Compartir](#)

E

Me gusta la idea 4, y me parecen útiles los personajes. Escribe un cuento utilizando la idea 4 para la trama y que utilice a los 7 personajes que creaste y la relación entre personajes. Debe contar con una introducción, un desarrollo, un clímax y una resolución. Divídalo en varios capítulos breves. La longitud final del cuento debe estar entre las 1200 y 1500 palabras.



Título: El Ritual de Medianoche

Capítulo 1: El Juego Comienza

Eva observaba la casa rural desde la entrada. La madera desgastada y los ventanales rotos eran el escenario perfecto para su experimento. Había reunido a sus amigos para un fin de semana diferente: invocar espíritus. No es que ella creyera en lo sobrenatural, pero le interesaba ver cómo sus amigos reaccionarían ante el miedo. Era psicóloga y quería probar sus límites.



Envía un mensaje a ChatGPT



ChatGPT puede cometer errores. Comprueba la información importante.



Captura 6: prompt final y comienzo del cuento



Anexo B. Cuento generado con ChatGPT4: “El Ritual de Medianoche”¹⁷⁵

Título: El Ritual de Medianoche

Capítulo 1: El Juego Comienza

Eva observaba la casa rural desde la entrada. La madera desgastada y los ventanales rotos eran el escenario perfecto para su experimento. Había reunido a sus amigos para un fin de semana diferente: invocar espíritus. No es que ella creyera en lo sobrenatural, pero le interesaba ver cómo sus amigos reaccionarían ante el miedo. Era psicóloga y quería poner a prueba sus límites.

—¡Vamos, gente! — exclamó Lucas, su típico tono sarcástico resonando en la sala. — ¿Cuándo empezamos a hablar con los muertos?

Ana, nerviosa y temblorosa, observaba las paredes como si algo pudiera saltar de ellas en cualquier momento. —No bromeen. No deberíamos estar jugando con estas cosas — dijo, su voz apenas un susurro.

—Es solo un juego — intervino María, la periodista, mientras revisaba algunas notas que había traído sobre la casa. —Aunque, según lo que he investigado, aquí vivía una familia que desapareció en circunstancias misteriosas hace más de cien años. Dicen que sus espíritus siguen atrapados aquí.

—¡Vaya! — exclamó Carlos, mejor amigo de María. —¿No es eso lo que siempre se dice de estas casas antiguas?

Sergio, el abogado, se mantenía en silencio. Sentado en la esquina, observaba cómo se disponían las velas para el ritual. La expresión en su rostro era inescrutable, como siempre.

¹⁷⁵ Se incluye el cuento tal como fue hecho por la IAG. Únicamente se añadieron modificaciones de formato para adaptarlo al de esta tesis.



Eva colocó el tablero de ouija en el centro y apagó las luces, dejando solo el parpadeo de las velas para iluminar la sala.

—¿Listos para empezar?— preguntó Eva, y todos asintieron, aunque el ambiente había cambiado. La broma de Lucas se desvanecía en el aire espeso y frío que ahora llenaba la casa.

Capítulo 2: La Primera Señal

El silencio era abrumador mientras Eva recitaba las palabras del ritual. El planchette del tablero de ouija no se movía. Todos mantenían las manos encima, esperando algo... cualquier cosa.

De repente, un golpe seco resonó en la casa. Ana gritó y retiró las manos de inmediato. —¡No quiero seguir con esto!— exclamó, su respiración agitada.

Lucas rió nerviosamente. —Debe ser el viento. Tranquila, Ana.

Pero algo había cambiado. Las velas parpadearon, y una ráfaga de aire frío recorrió la sala. Justo en ese momento, las luces se apagaron por completo. Se escucharon pasos rápidos y el sonido de un golpe fuerte. Cuando las luces volvieron, todos vieron a María en el suelo, inmóvil.

—¡María!— gritó Carlos, arrodillándose junto a ella. La sacudió, pero no respondió.

—Está muerta— murmuró Eva, en shock.

—Esto no puede ser real— dijo Lucas, retrocediendo con las manos en la cabeza.

Ana lloraba desconsoladamente, convencida de que habían despertado a un espíritu maligno. —Nos va a matar a todos— sollozaba.

Sergio permanecía en el fondo, observando en silencio.



Capítulo 3: El Detective

Horas más tarde, el inspector Marcos Hernández llegó a la casa. El detective, de mirada aguda y voz tranquila, caminó por la escena con calma. Había escuchado las versiones de los testigos, todos afirmaban que algo sobrenatural había causado la muerte. Pero Marcos no creía en fantasmas.

—Cuéntenme qué pasó— pidió mientras tomaba notas. Cada uno dio su versión. Ana estaba aterrada, convencida de que un espíritu los acechaba. Lucas intentaba restarle importancia, pero no podía ocultar su nerviosismo. Eva estaba frustrada, el experimento había salido terriblemente mal, y Carlos estaba devastado por la muerte de su amiga. Sergio, como siempre, era el más frío.

Marcos empezó a observar los detalles. La posición del cuerpo, las marcas en el cuello de María, la falta de señales de lucha. Sabía que algo no encajaba.

—Hay algo que no cuadra— murmuró para sí mismo.

Capítulo 4: Secretos Revelados

Al día siguiente, Marcos entrevistó a cada uno por separado. Eva explicó que el juego era solo un experimento psicológico, pero su frustración con la situación era evidente. Carlos parecía más preocupado por la muerte de María que por la idea de un fantasma, y su amistad con ella resultaba ser más compleja de lo que había pensado inicialmente.

Lucas admitió que no se había tomado nada en serio, pero no pudo ocultar lo nervioso que estaba desde la muerte. Ana seguía insistiendo en que un espíritu había hecho todo, y su miedo parecía sincero. Sin embargo, fue Sergio quien atrajo más la atención de Marcos. Era demasiado frío, demasiado calculador. Algo en él no cuadraba.



—¿Qué opinas de lo que pasó?— le preguntó Marcos.

Sergio lo miró con indiferencia. —No creo en fantasmas, inspector.

—Ni yo— respondió Marcos con una sonrisa.

Capítulo 5: La Revelación

Después de largas horas de investigación, Marcos comenzó a atar cabos. Descubrió que María había estado investigando algo más que la historia de la casa. Había encontrado información comprometedora sobre Sergio, relacionada con un caso de fraude en el que estaba involucrado.

Esa fue la clave. María había descubierto algo que Sergio no podía permitir que saliera a la luz. Aprovechó el caos del juego, la oscuridad y la superstición para cometer el asesinato sin levantar sospechas.

—Sergio— dijo Marcos cuando reunió a todos en la sala. —Sé que fuiste tú quien mató a María.

El grupo lo miró, atónito.

—¿Qué?— murmuró Carlos. —¿Por qué haría eso?

—María descubrió algo sobre él. Algo que habría arruinado su carrera. Sabía demasiado— explicó Marcos.

Sergio, finalmente acorralado, intentó mantener su calma, pero la tensión en sus hombros lo traicionaba. —No tienes pruebas.

—Claro que las tengo. Las marcas en el cuello de María indican que fue estrangulada. Las luces apagadas y la confusión te dieron la oportunidad perfecta. Utilizaste el juego como tapadera— concluyó Marcos.



Sergio se quedó en silencio. No había fantasmas. Solo un asesino metódico que había intentado usar el miedo para ocultar su crimen.

Capítulo 6: Justicia

Sergio fue arrestado, y aunque el grupo estaba commocionado por la verdad, también sintieron alivio. El miedo a lo sobrenatural se desvaneció con la confesión del asesino, pero las secuelas del juego mortal quedaron en cada uno de ellos.

Eva, con la culpa de haber iniciado el juego que desencadenó todo, juró no volver a mezclar la mente humana con lo inexplicable. Carlos, devastado por la pérdida de su amiga, se alejó del grupo. Ana aún seguía viendo sombras en la casa, aunque ahora sabía que el único monstruo era humano.

El inspector Marcos Hernández, satisfecho con la resolución del caso, dejó la casa rural con la firme convicción de que, aunque la oscuridad y el miedo a veces parecen invencibles, la verdad siempre encuentra su camino.

Fin



*Winternute era la mente colectiva,
tomaba las decisiones,
influenciando el cambio en el mundo exterior.
Neuromante era la personalidad.
Neuromante era la inmortalidad.
Marie-France debió haber implantado algo en Winternute,
la compulsión que lo impulsó a liberarse,
a unirse con Neuromante.*

Neuromante, William Gibson, 1984