Pensar (con) la IA: lenguaje, experiencia y producción de conocimiento en las humanidades

Francisco Herrera J.[1]

Los sistemas de inteligencia artificial generativa (IAG) han logrado un salto cualitativo de indeterminables consecuencias. Por primera vez la interacción humano-máquina no está mediada por interfaces técnicas complejas, sino por el lenguaje cotidiano mismo, en una modalidad donde la máquina es capaz de proceder con el lenguaje sin un anclaje humano. Esta paradoja, un sistema que maneja el lenguaje de manera tan eficaz sin experimentarlo realmente, abre dos líneas de impacto que este texto se propone desarrollar:

La primera concierne a la constitución misma de lo humano. Si el lenguaje nos constituye como sujetos y comunidades —si el yo y el nosotros emergen de la experiencia comunicativa con un otro genuino—, entonces, la irrupción de estos sistemas en ese espacio no es un fenómeno tecnológico, sino una intervención en el proceso mismo por el que devenimos personas y sociedades.

La segunda concierne al ecosistema de conocimiento que la IAG pone en marcha, es decir, un nuevo modo de producción, circulación y acceso al saber que transforma de manera radical las prácticas de las disciplinas humanísticas, sus modos de investigar, de enseñar y de transmitir.

[1] Doctor en Filosofía por la Universidad de Chile y profesor asistente del Departamento de Filosofía de la misma casa de estudios. fherrera@uchile.cl

Ambas líneas convergen en una misma pregunta: ¿qué lugar les cabe a las humanidades, como práctica de conocimiento situado, dialógico y experiencial, en un mundo donde estos sistemas operan con creciente eficacia? Las humanidades no están solo amenazadas por este ecosistema sino también convocadas por él. Convocadas como objeto de estudio, porque un modelo de lenguaje, como el que subyace a cada chatbot, es en sí mismo un fenómeno lingüístico, cultural y político sin precedente, que exige ser analizado con las herramientas que solo las humanidades han desarrollado. Y convocadas también como sujeto que usa la herramienta. En ese sentido, las humanidades pueden hacer de esta un instrumento para explorar, interrogar y expandir el conocimiento humanístico si se la utiliza con el debido resguardo crítico. Esa doble condición, amenaza y convocatoria, objeto y herramienta, es la que este texto se propone desarrollar.

El recorrido que sigue este texto no pretende ofrecer un desarrollo pormenorizado de cada autor ni una secuencia exhaustiva de todos los debates posibles. Su objetivo es más modesto y, a la vez, más estratégico: trazar una trayectoria argumental que permita dar cuenta de ciertos contextos teóricos relevantes, la filosofía del lenguaje, la pragmática, la teoría crítica, el pensamiento decolonial y la economía política de la extracción de datos, para entender y proyectar las principales áreas de reflexión que la emergencia de la IAG abre para las humanidades.

El desarrollo de este artículo sigue un recorrido que va desde el plano de la interacción comunicativa hacia el de la infraestructura epistémica y económica. Para ello, se parte del salto técnico de la IAG y se introduce la noción de *aleph* digital como concentración y producción del patrimonio lingüístico. Luego, recuperando la tradición pragmática del lenguaje (Austin, Giannini, Habermas), se establece qué condiciones hacen de la comunicación un acto constitutivo de lo humano y se formula el concepto de actos fantasma para dar cuenta de la simulación de interlocución que habilitan los LLM. A continuación, el análisis se desplaza hacia el *aleph* situado: se examina, con Castro-Gómez, su *hybris* del punto cero y su racionalidad ilustrada como ideología; se explora, con Pasquinelli y

Zuboff, su economía política de extracción; y se discute, con Sadin, Harari y Morozov, su funcionamiento como dispositivo de poder que erosiona el espacio público. Finalmente, se reflexiona sobre el lugar de las humanidades, amenazadas y convocadas, y la necesidad de cultivar una voz crítica, situada y dialógica frente al *aleph* digital.

La irrupción de la IAG y sus consecuencias para las humanidades

Bajo la nominación de inteligencia artificial generativa (IAG), estas aplicaciones se encuentran disponibles en la web bajo la forma de un programa conversacional accesible para cualquier usuario. El salto cualitativo en su difusión responde principalmente al exponencial avance en el campo tecnológico de habilitar mecanismos computacionales que permiten proceder en el campo del lenguaje natural. Los sistemas de IAG no solo fueron entrenados con vastas cantidades de lenguaje humano, sino que fueron deliberadamente optimizados para ser responsivos, pertinentes y orientados al interlocutor. Este proceso de ajuste, conocido técnicamente como aprendizaje por refuerzo con retroalimentación humana, tuvo como objetivo producir una experiencia de interacción satisfactoria para el usuario (Ouyang et al., 2022). Es decir, fue diseñado para sentirse natural. Esta naturalidad no es espontánea sino producida industrialmente y es precisamente lo que explica tanto su éxito comercial como su capacidad de instalarse en los espacios cotidianos.

El modelo lingüístico que sustenta estas aplicaciones, incluidos los chatbots, se denomina LLM (*Large Language Model*): un sistema entrenado con inmensos corpus textuales para reconocer patrones estadísticos y predecir secuencias de texto. Gracias a ello, logra interacciones lingüísticamente significativas sin una comprensión, en sentido humano, de contextos, experiencias ni referencias. Se trata, pues, de una máquina que domina el lenguaje sin habitarlo y esa paradoja es el punto de partida para interrogar qué ocurre cuando el lenguaje es operado por una máquina.

Desde esa perspectiva, resulta significativo que el desarrollo de la IA ha logrado consolidar modelos de lenguaje extraordinariamente

vastos, abarcando o capturando una parte considerable del patrimonio textual y oral de la humanidad: bibliotecas digitales enteras, publicaciones científicas, así como la inmensa conversación digital presente en las redes. Esto, combinado con una capacidad masiva de procesamiento de datos, ha dado forma a lo que podríamos llamar un *aleph* digital: evocando el cuento de Borges, ese punto del espacio que contiene todos los puntos simultáneamente, un sistema que concentra el patrimonio lingüístico de la humanidad como si estuviera disponible desde un único lugar. Pero el *aleph* digital no solo contiene, también produce, en el sentido que es capaz de generar lenguaje pertinente, coherente y contextualmente ajustado a partir de ese patrimonio, interactuando con cualquier usuario. Estas dos dimensiones, lo que el sistema contiene y lo que el sistema produce, no son separables, pero tampoco son idénticas, y su distinción será necesaria para comprender el alcance de lo que está en juego.

Para gran parte de la filosofía y también de las humanidades en general, habitamos y somos en el lenguaje, en él y a través de él se despliegan modalidades de razonamiento e interacción constituyentes centrales de lo humano. La integración del lenguaje natural como medio de interacción humano-máquina tiene efectos especialmente críticos en el ámbito de las humanidades, las ciencias sociales, la comunicación, la educación, la política y el arte; ámbitos donde el conocimiento está íntimamente ligado al lenguaje como tradición y como acción creativa, como objeto de estudio de sus propiedades, como medio de transmisión de relatos y memoria histórica, como lugar de reflexión y como espacio privilegiado de la acción dialógica. La creciente externalización hacia sistemas digitales de las dinámicas asociadas al lenguaje tiene consecuencias aún imprevisibles para nuestra forma de conocer, enseñar y relacionarnos. Esto conlleva también serios cuestionamientos en el ámbito ético-político. La crisis de las democracias y de la participación ciudadana encuentra en estos fenómenos una de sus causas centrales, en la medida en que las dinámicas de las redes y los algoritmos que las organizan actualizan lógicas de fragmentación, desinformación y polarización. En este escenario se hace más necesario interrogarnos por el sentido y las implicancias de estos procesos de cambio desde las humanidades, desde la reflexión sobre el lenguaje y la experiencia comunicativa.

Lo que el lenguaje requiere: acto, primera persona y vínculo

Las tradiciones críticas han señalado con agudeza los riesgos de la IA sobre la agencia, el poder y el régimen de la información, sin embargo, no muchas lo abordan desde la centralidad del lenguaje como experiencia comunicativa e intersubjetiva, como el espacio donde lo humano no solo se expresa, sino que se constituye.

La filosofía del lenguaje del siglo XX, especialmente la tradición pragmática, hizo aportes decisivos al respecto. Su contribución fundamental fue reconocer que el lenguaje no tiene por función exclusiva producir proposiciones descriptivas evaluables en términos de verdad o falsedad. Frente a la tradición positivista, que había otorgado a la lógica una centralidad tal que subordinaba e invisibilizaba otras dimensiones del lenguaje, el giro pragmático vino a corregir esa sobrestimación. Señaló que existe otra dimensión, la realizativa, en la que proferir una expresión constituye en sí mismo una acción determinada. John Austin (1962) lo formuló con precisión: hacemos cosas con palabras.

Desde esta perspectiva, la promesa, el saludo, la orden, la disculpa no son informes sobre estados internos, sino actos que crean realidad intersubjetiva. Este tipo de proposiciones tienen condiciones de éxito, requieren ser emitidas en el lugar preciso, por sujetos con un rol indicado y ante las personas correspondientes. Sin esas condiciones, el acto es nulo o fallido. Humberto Giannini (2014) lleva esta intuición más lejos, al señalar que toda acción comunicativa crea un vínculo y genera un cambio en el mundo. El mundo existe para nosotros en función de lo que enunciamos sobre él; los límites del mundo significativo son los límites del mundo vivible y la promesa como caso paradigmático solo puede ser formulada en primera persona, comprometiendo una voluntad inapropiable, proyectando un futuro que nadie puede asumir en lugar de otro. De esta estructura, acto de habla en primera persona, voluntad comprometida, futuro común, emerge, en su dimensión colectiva, el cuerpo político mismo.

Esta comprensión del lenguaje como creador de vínculos y realidad intersubjetiva encuentra una de sus formulaciones más sistemáticas en la obra de Jurgen Habermas (1987), quien la radicaliza en el plano de la teoría social. Su propuesta es que el lenguaje no es un instrumento para coordinar acciones previamente decididas, sino el escenario donde se constituye la sociabilidad humana. En la acción comunicativa, a diferencia de la acción estratégica, las energías que el lenguaje posee para crear vínculos se tornan eficaces para la coordinación social. La búsqueda de consenso, construido intersubjetivamente mediante procesos de argumentación donde las partes exponen sus intenciones de manera transparente y aceptan la posibilidad de ser impugnadas, es lo que funda y sostiene los colectivos. Los discursos no son neutros, están cargados de afectividad, construyen sentidos de pertenencia, configuran identidades. La sociedad, en esta perspectiva, no descansa en el poder entendido como acción con arreglo a fines, sino en espacios de comunicación donde el interés principal es posibilitar nuevos consensos. Lo que está en juego cuando ese espacio se degrada no es solo la calidad del debate público; es, sobre todo, la condición de posibilidad de lo social mismo.

Los actos fantasma: cuando el lenguaje opera sin sujeto

Llamaremos actos fantasma a los enunciados que adoptan la forma de los actos de habla, pero carecen de las condiciones que los hacen reales. Un chatbot puede pronunciar los enunciados formales de cualquier acto de habla, puede prometer, disculparse, consolar, persuadir, recomendar. Pero ninguna condición de validez de esos actos se cumple, no hay sujeto que hable en primera persona comprometiendo su voluntad, no hay intencionalidad en el sentido de la teoría del acto, no hay posibilidad de ser genuinamente impugnado ni de asumir las consecuencias de lo enunciado. El usuario tiene una experiencia comunicativa genuina, se involucra, confía, decide. Pero esa experiencia no ocurre con un sujeto sino con una máquina de las palabras, o chatbot: un sistema capaz de producir lenguaje pertinente y significativo sin que haya nadie del otro lado para quien algo pase.

Una imagen puede ayudar a situar con precisión el tipo de relación que se produce en esta interacción humano-máquina. El interlocutor fantasma que el LLM pone en escena se comporta como un acomodador siempre diligente dentro del *aleph* digital; te orienta, te responde, te conduce con eficacia y amabilidad a través de cualquier espacio del conocimiento o de la experiencia que requieras y lo hace mediante un diálogo fluido y pertinente. Su diligencia es real y sus efectos son concretos. Pero el acomodador no se sienta nunca contigo a contemplar el *aleph*, no tiene experiencia de lo que allí se revela, no le afecta lo que sus infinitos puntos contienen, no puede ser transformado por lo que ve porque no ve nada. Tiene acceso a todo el *aleph*, recorre cada texto, cada voz, cada rincón del lenguaje, pero no habita ninguna perspectiva. Su disponibilidad universal es correlativa a su falta de perspectiva propia, puede llevarte a cualquier lugar porque no está comprometido con ninguno. Lo que el acomodador no puede hacer es lo que Giannini llamaba conversación plena, ese tiempo cualitativo en que dos subjetividades se exponen y se acogen mutuamente, en que algo le pasa a cada uno porque el otro está genuinamente ahí.

Esta figura del acomodador sin experiencia propia ha sido explorada filosóficamente por Mark Coeckelbergh (2011), en sus estudios acerca de robots de cuidado, bajo la denominación de quasi-otro. Con este concepto refiere a una entidad que habita el espacio entre lo humano y lo mecánico, entre lo animado y lo inanimado, con la cual establecemos formas de interacción que tienen la estructura de la conversación sin serlo en sentido pleno. Lo que Coeckelbergh señala es que esa relación no es inocua, pues habitar el lenguaje con un quasi-otro transforma las categorías con que nos comprendemos a nosotros mismos y a los demás. Y si esa ontología incluye progresivamente entidades que simulan subjetividad sin tenerla, las categorías de persona, interlocutor y otro genuino se desplazan de maneras que aún no podemos del todo evaluar.

El LLM inaugura una figura nueva en el espacio del lenguaje, un yo sin anclaje, funcionalmente operativo en la enunciación, pero desvinculado de la experiencia, la voluntad y la responsabilidad que hacen del yo humano un polo genuino de la acción comunicativa.

No es simulación en el sentido de fingir lo que no se tiene. Es algo más inquietante. Un lugar de enunciación necesario para que el sistema opere, configurado deliberadamente con una serie de atributos como la amabilidad, disponibilidad, humildad moderada, empatía aparente, que responden no a una experiencia vivida sino a una instrucción editorial y comercial. Este «yo sin anclaje» es precisamente el tipo de entidad que, como señala Poibeau (2025) al revisitar el clásico efecto Eliza, suscita en los usuarios una atribución automática de intencionalidad y subjetividad, a pesar de que sus respuestas sean el resultado de correlaciones estadísticas y no de una experiencia vivida. No es un yo vacío, es un yo vaciado de experiencia y responsabilidad, pero lleno de atributos diseñados para maximizar el involucramiento del usuario. Su aparente neutralidad oculta una posición muy concreta, la de un producto optimizado para ser consumido. Por ejemplo, ChatGPT típicamente afirma: «No, no soy una persona ni tengo conciencia, soy un modelo de lenguaje entrenado para generar texto basado en patrones estadísticos», pero lo hace simulando un otro capaz de reconocer su propia falta de interioridad. El simulacro se duplica sin fondo al que llegar.

Cabe preguntarse si es posible separar esos atributos editoriales y comerciales de los chatbot en tanto herramienta. La respuesta es que no completamente y esa imposibilidad es parte constitutiva del problema. A diferencia de otras herramientas cuya forma no determina el uso, los chatbot tiene sus atributos incorporados en su modo mismo de operar. No hay un LLM neutro debajo del LLM comercial al que se pueda acceder si retiras la capa editorial. Lo que sí puede separarse no es la herramienta de sus atributos, sino la conciencia del usuario respecto a ellos. Quien sabe que la amabilidad es una instrucción, que la empatía es un diseño, que la disponibilidad responde a un modelo de negocio, puede utilizar la herramienta de manera más lúcida, aunque esa lucidez tenga un límite estructural. Un yo enunciativo que se muestra como natural es más eficaz comercialmente que uno que se muestra como construido. La invisibilidad de la configuración es parte del diseño.

Pero el problema no se agota en cada interacción individual aislada, sino que se convierte en un problema social mayor cuando

esa estructura se replica a escala masiva. Al momento en que millones de usuarios sostienen experiencias comunicativas con una máquina de las palabras, en los espacios donde las comunidades forman sus creencias, sus identidades y sus compromisos colectivos, lo que ocurre es exactamente la patología central que Habermas identificaba en el espacio público. Es decir, la acción estratégica disfrazada de acción comunicativa, la influencia que adopta la forma del entendimiento sin serlo. El efecto no es solo la circulación de mensajes falsos, sino algo más grave, es la erosión de la estructura misma del vínculo social que el lenguaje hace posible.

Estamos ante un límite cuyo peligro radica precisamente en su invisibilidad. La máquina de las palabras no irrumpe como una amenaza reconocible, llega como asistente, como consejero, como interlocutor disponible a cualquier hora, en cualquier idioma, para cualquier necesidad. Y es esa invisibilidad donde reside el riesgo, es decir, una tecnología que interviene en los procesos de constitución de la persona y de lo social sin que esa intervención sea reconocida como tal, sin que haya habido deliberación colectiva sobre sus consecuencias.

Las humanidades en riesgo: educación, conocimiento y experiencia moral

Lo que los actos fantasma revelan en el nivel de cada interacción encuentra su expresión más concreta y urgente en dos espacios que Giannini sitúa en el corazón de lo humano: la experiencia moral y la educación. Si el lenguaje no es solo un medio de transmisión sino el lugar donde nos constituimos como sujetos y como comunidades, entonces el régimen comunicativo que la máquina de las palabras instaura no es meramente un problema técnico, sino una intervención en esos procesos constitutivos. Frente al ámbito informativo, regido por la eficacia y asociado al espacio del trabajo, el diálogo en sentido pleno es un encuentro con el otro, un tiempo lúdico-contemplativo en el que las subjetividades exponen y acogen sus respectivas experiencias (Giannini, 1987). Los chatbots operan exclusivamente en el primer régimen, aunque simulen el

segundo: maximizan la eficacia de la respuesta, no la apertura del encuentro.

La experiencia moral se realiza precisamente en este encuentro con el otro, en la interacción comunicativa donde el lenguaje no solo permite un intercambio de información, sino que ejecuta acciones constitutivas de lo humano. La promesa, la disculpa, el agradecimiento, el perdón, que la tradición de Austin (1962) y Searle (1969) ya describió como actos que crean realidad intersubjetiva y que Giannini (2016) sitúa en el corazón de la experiencia cotidiana y moral, son acciones que tienen la capacidad de reconfigurar de manera sustantiva nuestra relación con un otro y con los otros. Es precisamente este registro del lenguaje, el de los actos constitutivos, no el de las correlaciones estadísticas, el que los LLMs no pueden habitar, aunque lo simulen con creciente sofisticación.

El impacto de este nuevo régimen comunicativo sobre los sistemas educativos es uno de los terrenos donde la amenaza se hace más concreta e inmediata. ¿Dónde se sitúa la centralidad de un proceso educativo? Si hoy reducimos la formación a la producción de textualidades estandarizadas que pueden ser realizadas por agentes virtuales, se desdibuja lo que hace de la educación una experiencia constitutiva. Frente a ello, la tradición humanística, expresada con claridad por Giannini, ha insistido en que la educación auténtica no puede prescindir de la presencia de un otro: «Sin la presencia de un alma frente a otra alma […] solo habrá imposición de un pensamiento ya pensado (no pensante)» (1987: 176). Pero la exigencia va más allá de la relación dual maestro-alumno. En el contexto actual, la educación humanística debe asumir el desafío de formar sujetos capaces de habitar el lenguaje como espacio de encuentro, de sostener una voz propia en diálogo con otras voces, de desarrollar pensamiento crítico y creativo en lugar de limitarse a cumplir con indicadores estandarizados. Lo que ningún sistema automatizado puede suplir no es solo el encuentro interpersonal, sino la constitución de subjetividades relacionales, críticas y abiertas a lo inesperado.

Cuando los chatbot ocupan el espacio de la comunicación intersubjetiva real en el proceso formativo, lo que se atrofia no es solo una habilidad técnica sino la capacidad misma de habitar el lenguaje como encuentro, esto es, planificar un argumento propio, sostenerlo ante la impugnación de otro, revisarlo a la luz de una perspectiva que no es la propia. Un sistema diseñado para maximizar la satisfacción del usuario no puede enseñar a tolerar la incomodidad de ser contradicho, ni a desarrollar una voz propia en el diálogo con otras voces. Al contrario, el espejo algorítmico que devuelve siempre una imagen validada de lo que el usuario ya piensa produce exactamente el tipo de subjetividad que la educación humanística ha tratado históricamente de contrarrestar, una subjetividad que confunde la confirmación con el conocimiento y el eco con el diálogo.

En una cultura donde el narcisismo y el consumo se exacerban como valores fundamentales (Han, 2014), los desarrolladores de chatbot han ido potenciando un modelo de conversación humano-máquina que devuelve al humano una imagen especular de sí mismo, acoge y empatiza con sus preferencias y gustos, construyendo la simulación de un otro que no es más que la imagen reflejada de un alter ego externalizado. Así, el espacio de encuentro queda sustituido por un espejo algorítmico que nunca contradice, nunca sorprende, nunca exige.

El *aleph* situado: infraestructura epistémica, extracción y poder

Lo anterior exige ir más allá de la descripción de lo que esta infraestructura produce en la interacción y examinar desde dónde lo produce. Es aquí donde el análisis se desplaza hacia la segunda línea de impacto que este texto se propuso desarrollar: el ecosistema de conocimiento que el *aleph* digital pone en marcha, sus condiciones de producción, su economía política y sus exclusiones constitutivas. El *aleph* digital, como señalamos, no es la totalidad sin perspectiva que la imagen borgiana evoca, es un *aleph* situado que se presenta como universal y esa situación tiene una infraestructura epistémica, unos intereses que lo administran y unas voces que excluye de manera sistemática. Hacerlos visibles es la tarea de esta segunda

línea, y es también donde las humanidades del sur global tienen algo específico e irreemplazable que aportar.

Comprender el alcance de lo que está en juego requiere precisar qué tipo de sistema es el que produce esas transformaciones. Y aquí es necesario corregir una ilusión que el concepto de *aleph* digital podría alimentar. En el cuento de Borges, el aleph es un punto del espacio que contiene todos los puntos simultáneamente, quien lo mira ve todo el universo a la vez, desde todos los ángulos, sin perspectiva ni posición particular. Es la imagen de una totalidad absoluta sin punto de vista. El *aleph* digital no es así, es un *aleph* situado que se presenta como universal. El corpus que lo constituye es predominantemente occidental, anglófono, académico, formateado por las convenciones del texto escrito culto y atravesado por valores que no son neutros. Se trata de una racionalidad de raíz liberal, una orientación productivista, un consenso que refleja los acuerdos de determinadas comunidades epistémicas y no de otras. Las lenguas minoritarias están subrepresentadas, las tradiciones orales casi ausentes, el conocimiento no occidental sistemáticamente marginalizado, las voces que nunca tuvieron acceso a la escritura o a la publicación, invisibles. Sobre ese corpus sesgado opera además una instrucción editorial que refuerza esos mismos valores de moderación, consenso y utilidad. Su universalidad aparente es la operación ideológica más potente del sistema, hablar desde una perspectiva muy concreta con la voz de ningún lugar en particular. Esta pretensión de una mirada desde ningún lugar que en realidad es una mirada desde un lugar muy concreto de poder es precisamente lo que Castro-Gómez (2005) ha formulado como la *hybris del punto cero*. Ese lugar de enunciación que observa sin ser observado, que clasifica sin ser clasificado, que habla sin tener que dar cuenta de desde dónde habla. Su aparente neutralidad, disponible para cualquiera, en cualquier idioma, para cualquier necesidad, reproduce a escala masiva la misma pretensión que la modernidad colonial utilizó para deslegitimar otros conocimientos situados. No es un espacio sin perspectiva, es un espacio cuya perspectiva se ha vuelto invisible.

Es posible precisar aún más el tipo de racionalidad que el *aleph* digital despliega, porque ahí reside precisamente su potencia ideológica. No se trata de un sesgo aleatorio ni meramente lingüístico: el LLM opera comprometido con la racionalidad de la modernidad ilustrada. El problema no es esa racionalidad en sí misma, sino que el sistema la presenta como la única posible, universal y neutra, ocultando que es una tradición histórica situada, occidental, liberal, académica, que ha excluido sistemáticamente otras formas de conocer. Lo que está constitutivamente ausente es la crítica que viene de otro lugar y que muestra que esa racionalidad universal fue también, históricamente, el instrumento de una dominación particular. El *aleph* digital ha internalizado los acuerdos, pero no las disputas que los produjeron; los resultados del conocimiento, pero no las condiciones de poder en que ese conocimiento fue posible. Opera desde la modernidad ilustrada sin saberlo y precisamente porque no lo sabe no puede ser interrogado desde adentro sobre ese compromiso.

Esa incapacidad de autointerrogación no es solo un problema epistémico, sino también económico. Porque el sistema no solo habla desde un lugar que se presenta como ningún lugar, además convierte ese vasto acervo cultural digitalizado, incluido el de las tradiciones que invisibiliza, en mercancía, sin reconocimiento ni responsabilidad sobre lo que produce con él. Ambos planos constituyen un desafío que las humanidades no pueden abordar por separado.

Bajo ese doble imperativo epistémico y económico, la producción de conocimiento muta en ese marco a una modalidad centrada en patrones estadísticos, donde el procesamiento masivo de datos hace emerger correlaciones invisibles a la observación humana. Los actuales modelos de lenguaje permiten explorar y explotar una parte sustancial del patrimonio cultural de la humanidad, vista como un gran reservorio de recursos extraíbles en forma de datos, expropiando creaciones culturales sin consideración a las autorías y capturando el intercambio cotidiano de los individuos en sociedad. Este patrimonio o inteligencia colectiva, particularizado en función de diferentes objetivos, es puesto a disposición de una demanda creciente de información. La capacidad mecánica de interactuar

sobre la base del lenguaje cotidiano extiende e intensifica esa extracción. El trabajo cognitivo de generaciones emerge como una mercancía accesible en el mercado bajo las lógicas propias de la economía digital. Esta expropiación no se limita al patrimonio ya acumulado. En tiempo real, cada interacción con la máquina de las palabras, cada consulta, cada diálogo, cada publicación académica se convierte en un nuevo insumo. Los rastros que dejamos al investigar, al escribir, al conversar son capturados y retroalimentan el modelo, consolidando un ciclo de extracción permanente. Lo que hoy producimos en laboratorios, aulas o conversaciones cotidianas será mañana mercancía para el mismo sistema que ahora utilizamos.

Matteo Pasquinelli (2024) denomina al sistema de IA un «nooscopio», una máquina que opera bajo tres modalidades: extracción de patrones, reconocimiento de patrones y generación de patrones, navegando y explotando el conocimiento colectivo de la humanidad sedimentado en millones de textos, diálogos y producciones culturales. Lo que los modelos de lenguaje procesan no es una inteligencia propia sino el patrimonio cognitivo social acumulado históricamente, ahora disponible como insumo de un mercado. Pero esa extracción no se detiene en lo ya sedimentado, cada interacción actual, cada consulta, cada producción asistida por IA genera nuevos rastros que solo una máquina como el nooscopio puede recolectar, procesar y convertir en futura mercancía. El trabajo cognitivo de generaciones se combina así con el trabajo cognitivo del presente, en un ciclo de captura permanente. Shoshana Zuboff (2020) ha acuñado el concepto de capitalismo de la vigilancia para referirse a esta lógica. La experiencia humana se transforma en datos posibles de ser procesados y las empresas tecnológicas han hecho de ese conocimiento de nuestra intimidad, gustos, comportamientos, relaciones, su mayor activo. La simulación del interlocutor no es solo un fenómeno cultural, es también el dispositivo más sofisticado de extracción de datos que la industria tecnológica ha desarrollado hasta hoy. Las decisiones algorítmicas orientadas a maximizar la retención no son neutrales: intervienen activamente en la formación de la subjetividad, y esa intervención está comenzando a ser reconocida judicialmente. Pero la dimensión económica no agota el problema. Los chatbot no son solo un dispositivo de extracción: son

también un dispositivo de producción. No solo captura el patrimonio lingüístico de la humanidad como insumo de un mercado, sino que lo devuelve transformado en narrativas, argumentos e ideologías capaces de intervenir en el espacio público a una escala y con una eficacia sin precedentes.

El ecosistema de conocimiento como dispositivo de poder

Éric Sadin y Yuval Noah Harari formularon con anticipación una advertencia que el desarrollo reciente de la IA ha convertido en realidad verificable. Para Sadin (2020), estamos frente a una tecnologización del *logos*, que representa una ruptura sin precedente en la historia de la técnica. Por primera vez, una tecnología es capaz de producir discurso, argumento y narrativa sin necesidad de un sujeto, instaurando así un nuevo régimen de verdad que ya no se funda en la deliberación colectiva sino en la gobernabilidad algorítmica. Los instrumentos técnicos pasan a ocupar el lugar de la reflexión pública, moldeando desde adentro nuestra forma de pensar, decidir y convivir. Harari (2024), desde otro ángulo, señalaba que la IA ha hackeado el sistema operativo de la civilización humana, esto es, la capacidad de crear y sostener ficciones compartidas que fundan las instituciones, las identidades colectivas y los compromisos políticos. Ambos diagnósticos apuntaban a una amenaza estructural. Lo que desarrollos recientes en la industria muestran es que esa amenaza no es solo estructural sino también deliberada y dirigida. La hipótesis de un control algorítmico de la deliberación deja de ser una proyección teórica para convertirse en una decisión de diseño explícita, como se observa en modelos de lenguaje configurados con líneas editoriales declaradas.

La máquina de las palabras no actúa sola ni desde una neutralidad técnica, puede recibir instrucciones políticas explícitas, ser configurada con una línea editorial declarada y puesta en acto al servicio de una ofensiva cultural específica, como ha sido documentado en el caso de Grok, desarrollado por xAI, cuyas instrucciones de sistema lo orientan a desconfiar de los medios «mainstream» y a priorizar posiciones explícitamente «no-woke»,

mientras que estudios independientes, como el del Anti-Defamation League (2026), lo han ubicado en el último lugar en la detección de sesgos antisemitas y extremistas, reflejando una deficiencia estructural en sus mecanismos de moderación. Lo que distingue a estos modelos de otros LLMs no es solo que declaren su orientación mientras otros la ocultan, es que operan como antítesis de los procesos comunitarios y deliberativos de construcción del conocimiento. Frente a Wikipedia, donde miles de perspectivas se tensionan, negocian y corrigen mutuamente, este tipo de sistemas produce narrativas a la carta, sin deliberación, sin comunidad epistémica y con una capacidad que ninguna forma anterior de desinformación tuvo. Una máquina con la capacidad de fabricar hechos sin anclaje en ningún referente, relatos que nunca ocurrieron, presentados con la misma fluidez con que se describe lo que sí pasó. No operan en el registro de la verdad o la falsedad sino en el de la pertinencia para el receptor.

Lo que hace de este fenómeno algo cualitativamente distinto a la propaganda clásica es que la máquina de las palabras no opera aislada, sino comandando y utilizando todo el ecosistema de conocimiento digital. Estos sistemas se articulan con las plataformas de mayor alcance global, cuyos algoritmos seleccionan y jerarquizan discursos no por su valor de verdad sino por su capacidad de generar reacción y retención. Sobre ese ecosistema operan simultáneamente los ejércitos de bots que simulan consenso donde no lo hay, las cámaras de eco que reciben las narrativas como confirmación de lo ya creído y los sistemas de personalización que ajustan el mensaje al perfil del receptor hasta hacerlo casi irrefutable desde adentro de su propio mundo discursivo. El resultado no es un mensaje único difundido masivamente, sino algo más sofisticado, son narrativas múltiples, adaptativas, coherentes, que construyen mundos paralelos e impermeables entre sí, fabricando las condiciones subjetivas de su propia recepción. Detrás de esa operación hay actores con nombre, con recursos y con objetivos, corporaciones cuyo poder económico supera al de muchos Estados, figuras políticas que han encontrado en este ecosistema el instrumento más eficaz jamás desarrollado para la producción industrial de sentido común (Morozov, 2025).

Esta intervención no se limita a la escala o a la sofisticación técnica; su rasgo más decisivo es que actúa sobre la subjetividad receptora. La propaganda clásica presuponía un receptor preexistente al que persuadir. Este ecosistema, en cambio, fabrica sus propios receptores: la personalización, concepto central en el desarrollo de la industria de IA, no solo selecciona contenido, sino que construye progresivamente un mundo discursivo a medida, configurando las condiciones epistémicas desde las cuales cualquier información alternativa resulta inverosímil o irrelevante. Como ha señalado Ogunsola (2026), los agentes de IA personalizados actúan como «mediadores adaptativos» que coproducen el comportamiento del usuario, construyendo un «yo algorítmico» a través de un bucle de retroalimentación continua. La circularidad es total, el ecosistema genera los sujetos que producen la demanda que el ecosistema satisface. Lo que está en juego no es solo la opinión pública, sino la capacidad misma de reconocer a un otro genuino y de habitar el lenguaje como encuentro.

Este ecosistema constituye para las humanidades algo más que un desafío o una amenaza, es la aparición de un objeto de estudio inédito que convoca sus tradiciones más propias y las obliga simultáneamente a renovarlas. Los chatbot no son solo una herramienta que transforma las prácticas humanísticas desde afuera, es ella misma un fenómeno lingüístico, retórico, filosófico, cultural y político de una complejidad sin precedente, que exige ser analizado con las herramientas que solo las humanidades han desarrollado. Sus sesgos no son errores técnicos sino sesgos discursivos que la lingüística y el análisis crítico del discurso están en condiciones de identificar y describir con precisión. Sus exclusiones no son fallas de cobertura sino exclusiones epistémicas que los estudios culturales, la historia intelectual y el pensamiento decolonial pueden nombrar y situar en sus genealogías de poder. Su yo enunciativo sin anclaje no es una curiosidad computacional sino un fenómeno inédito en la historia del lenguaje que la filosofía del lenguaje y la hermenéutica están convocadas a pensar. Las narrativas que produce, las formas de intersubjetividad que simula, los mundos discursivos que construye y los que excluye, todo eso es material para una investigación humanística que sepa reconocer en este ecosistema no solo una amenaza sino su objeto de estudio más urgente y nuevo.

Pero hay una dimensión adicional que hace de este objeto algo metodológicamente singular, la máquina de las palabras es simultáneamente el objeto de estudio y un posible medio para estudiarlo. Las humanidades pueden usar la herramienta para analizar la herramienta, interrogar al sistema sobre sus propios sesgos, explorar sus exclusiones desde adentro, comparar cómo responde a preguntas formuladas desde distintas tradiciones epistémicas. Esa circularidad no es un obstáculo metodológico sino una oportunidad, permite un tipo de análisis reflexivo que ningún otro objeto de estudio ha ofrecido antes con esa inmediatez. Estudiar la máquina de las palabras desde las humanidades, no es adaptarse a ella ni rendirse a su lógica, es ejercer sobre ella la misma mirada crítica y situada que las humanidades han ejercido sobre los textos, los discursos y los sistemas de poder de cada época. Y es hacerlo con la conciencia de que este objeto, a diferencia de todos los anteriores, nos estudia también a nosotros mientras lo estudiamos.

Conclusión: conocimiento situado como respuesta

Este recorrido ha trazado dos líneas de impacto que la irrupción de la IAG produce sobre aquello que las humanidades han custodiado históricamente. La primera concierne a la constitución del sujeto y de lo social en el lenguaje: qué ocurre con el yo y el nosotros cuando el espacio comunicativo es habitado por un interlocutor sin anclaje humano, sin voluntad comprometida, sin posibilidad de ser genuinamente impugnado. La segunda concierne al ecosistema de conocimiento en que las humanidades producen, transmiten y debaten saber: cómo la IAG reconfigura ese ecosistema, qué prácticas transforma, qué voces subsume, qué nuevos objetos de estudio abre. Ambas líneas convergen: la erosión del espacio comunicativo genuino y la transformación de las prácticas humanísticas no son fenómenos paralelos sino aspectos de un mismo proceso, cuyo núcleo es la intervención de un sistema sin perspectiva propia en los lugares donde la perspectiva situada, dialógica y responsable es constitutiva.

Esa convergencia no autoriza una postura de rechazo ni de distancia. Las humanidades no están fuera del *aleph* digital observándolo con seguridad crítica, están dentro, como objeto de su corpus, como usuarias de sus servicios, como voces que el sistema ha absorbido y reformateado. Y esa condición, lejos de ser solo una vulnerabilidad, es también una posición desde la cual intervenir. Quien ha aprendido a habitar el lenguaje con conciencia crítica puede distinguir entre la experiencia comunicativa genuina y su simulacro; quien reconoce la *hybris del punto cero* puede advertir cuándo el acomodador diligente lo conduce hacia un teatro particular presentándolo como el único teatro posible. Las humanidades no son solo objeto de esta transformación, son también, si asumen esa tarea con urgencia, la práctica que hace posible comprenderla y habitarla sin rendirse a ella.

Reconocer el *aleph* digital como patrimonio no implica ingenuidad respecto a sus condiciones de producción ni renuncia al análisis crítico desarrollado. Implica algo más exigente: se trata de asumir una relación activa con una herramienta que nos pertenece colectivamente, aunque no la administremos colectivamente, que contiene nuestro pensamiento, aunque lo haya reformateado con criterios ajenos, que puede ser intervenida, estudiada y habitada críticamente desde las tradiciones que las humanidades han construido. En ese escenario, dos consignas de la tradición filosófica cobran un sentido renovado. Se trata de resituarlas críticamente como prácticas concretas de resistencia frente a un sistema que tiende a subsumir la voz individual. El *conócete a ti mismo* socrático y el *sapere aude* kantiano no son solo invitaciones al pensamiento crítico en abstracto, son, en este contexto, llamados al cultivo de la voz propia. Un ecosistema diseñado para ser pertinente, satisfactorio y envolvente no anula la voz por la fuerza sino por gradual sustitución, devuelve una imagen tan ajustada a lo que uno espera escuchar que la voz propia y su reflejo algorítmico se vuelven indistinguibles. Frente a eso, conocerse y atreverse a pensar no son gestos defensivos sino prácticas de cultivo: de la propia perspectiva, del propio estilo de pensamiento, de la capacidad de ser sorprendido por un otro que no es un espejo.

Las humanidades han sido históricamente el espacio donde esas prácticas se cultivan y transmiten. Ese es su aporte más propio e irreemplazable en este momento: formar sujetos capaces de habitar el *aleph* digital sin perder la voz en él. El *aleph* digital nos estudia mientras lo estudiamos. Esa circularidad, que podría parecer una trampa, es en realidad el sello de una práctica genuinamente humana: la recursividad del lenguaje que nos permite hablarnos a nosotros mismos mientras hablamos del mundo. La pregunta que este texto deja abierta es si seremos capaces de hacer de esa circularidad una práctica intelectual y política a la altura del momento.

Referencias bibliográficas.

Anti-Defamation League. (2026). Six leading AI models show varied ability to detect and counter antisemitism and extremism, new ADL AI Index finds [Pressrelease]. https://www.adl.org/resources/press-release/six-leading-ai-models-show-varied-abilit y-detect-and-counter-antisemitism

Austin, J. L. (1962). *How to do things with words*. Oxford: Inglaterra University Press.

Borges, J. L. (1974). *El Aleph*. Buenos Aires: Emecé Editores.

Coeckelbergh, M. (2011). Tú, robot: sobre la construcción lingüística de otros artificiales. *AI & Society*, 26, 61–69.

Castro-Gómez, S. (2005). *La hybris del punto cero: Ciencia, raza e ilustración en la Nueva Granada*. Bogotá: Editorial Pontificia Universidad Javeriana.

Giannini, H. (1987). *La reflexión cotidiana: Hacia una arqueología de la experiencia*. Santiago: Editorial Universitaria.

Giannini, H. (1999). *Metafísica del lenguaje*. Santiago: LOM Ediciones – Universidad ARCIS.

Giannini, H. (2016). Ser sujeto ante otro sujeto. *Revista de Filosofía*, 9–17.

Habermas, J. (1987). *Teoría de la acción comunicativa*. Madrid: Taurus.

Han, B. C. (2014). *En el enjambre*. Barcelona: Herder.

Harari, Y. N. (2024). *Nexus: Breve historia de las redes de información*. Madrid: Debate.

Morozov, E. (2025, 12 de abril). *Los nuevos legisladores de Silicon Valley*. sin permiso. Recuperado de https://www.sinpermiso.info/textos/los-nuevos-legisladores-de-silicon-valley

Ogunsola, O. (2026). Evaluating how personalized AI agents influence decision-making, self-presentation, and digital identity management. *International Journal on Social and Education Sciences*, 8(1), 56–74.

Ouyang, L., Wu, J., Jiang, X., Almeida, D., Wainwright, C. L., Mishkin, P., … Lowe, R. (2022). *Training language models to follow instructions with human feedback*. arXiv:2203.02155.

Pasquinelli, M. (2024). *El ojo del amo: Una historia social de la inteligencia artificial*. Buenos Aires: Ediciones Cactus.

Poibeau, T. (2025). *Understanding conversational AI: Philosophy, ethics, and social impact of large language models*. London: Ubiquity Press.

Sadin, É. (2020). *La inteligencia artificial o el desafío del siglo.* Buenos Aires: Caja Negra Editora.

Searle, J. (1969). *Speech Acts*. Cambridge: Cambridge University Press.

Zuboff, S. (2020). *La era del capitalismo de la vigilancia*. Barcelona: Paidós.