

Artículos

Dilemas semióticos de la enunciación en ChatGPT

The Semiotic Dilemmas of Enunciation in ChatGPT Les dilemmes sémiotiques de l'énonciation dans ChatGPT

María Clara Lucifora

Universidad Nacional de Mar del Plata

mlucifora@mdp.edu.ar

Resumen

En este trabajo, pretendemos explorar el tipo de enunciación de ChatGPT y algunos dilemas semióticos asociados a esta inteligencia artificial. Realizamos esta indagación utilizando constructos teóricos que nos permiten analizar este fenómeno en desarrollo. Concluimos que, si bien se puede hablar de enunciación en estos casos, hay que distinguirla claramente de la enunciación humana, calificándola como “maquinica”. Esta denominación pone de manifiesto las múltiples diferencias en términos de producción de sentido, posición ideológica, rol del usuario/destinatario, efectos semióticos de la interacción, entre otros aspectos. Este tema no ha sido abordado sistemáticamente hasta el momento, dada su novedad y actualización permanente. Por ello, consideramos necesario construir marcos teóricos flexibles que propicien análisis diversos sobre el estatuto y los efectos concretos que supone incorporar estos modelos de lenguaje a nuestra existencia.

Palabras clave: ChatGPT, enunciación, modelo de lenguaje, producción de sentido, usuario

Abstract

In this work, we intend to explore the type of ChatGPT enunciation, and some semiotic dilemmas associated with it. We carry out this inquiry using theoretical constructs that allow us to analyze this developing phenomenon. We conclude that, although it is possible to speak of enunciation in these cases, it must be clearly distinguished from human enunciation, qualifying it as “machinic”. This designation highlights the multiple differences in terms of production of meaning, ideological position, role of the user/recipient, semiotic effects of interaction, among other aspects. This issue has not been systematically addressed so far, given its novelty and permanent updating. For this reason, we consider it necessary to build flexible theoretical frameworks that promote diverse analyses on the status and concrete effects of incorporating these language models into our existence.

Keywords: ChatGPT, enunciation, language model, production of meaning, user

Résumé

Dans ce travail, nous avons l'intention d'explorer le type d'énonciation du ChatGPT et certains dilemmes sémiotiques qui y sont associés. Nous réalisons cette recherche en utilisant des constructions théoriques qui nous permettent d'analyser ce phénomène en développement. Nous concluons que, s'il est possible de parler d'énonciation dans ces cas, celle-ci doit être clairement distinguée de l'énonciation humaine, en la qualifiant de « machinique ». Cette dénomination met en évidence les multiples différences en termes de production de sens, de position idéologique, de rôle de l'utilisateur/destinataire, d'effets sémiotiques de l'interaction, entre autres aspects. Ce thème n'a pas été abordé de manière systématique jusqu'à présent, compte tenu de sa nouveauté et de sa mise à jour permanente. C'est pourquoi nous considérons qu'il est nécessaire de construire des cadres théoriques flexibles favorisant différentes analyses sur le statut et les effets concrets concernant l'éventuelle incorporation de ces modèles linguistiques à notre existence.

Mots-clés : ChatGPT, énonciation, modèle de langage, production de sens, utilisateur

Introducción

ChatGPT es un software de inteligencia artificial, desarrollado por OpenAI, cuya web se habilitó el 30 de noviembre de 2022. En su primer mes, se registraron diez millones de usuarios por día, por lo que se convirtió en la tecnología más rápidamente adoptada en la historia. Nació de la convergencia de muchas otras tecnologías desarrolladas con anterioridad. Su versión inicial se basó en GPT-3 (*generative pre-trained transformer 3*), un modelo de lenguaje generativo que no sólo tiene la capacidad de conocer e identificar partes de lenguaje, sino también de predecir palabras y completar textos. Sin embargo, ChatGPT es una versión mejorada de ese modelo, pues también puede responder preguntas.

En este trabajo, nos proponemos explorar las características del tipo de enunciación que lleva a cabo ChatGPT y algunos dilemas semióticos asociados a ella. Realizaremos esta indagación utilizando ciertos constructos teóricos que nos permiten pensar en este fenómeno en desarrollo. Este tema no ha sido abordado sistemáticamente hasta el momento; por ello, y por el enorme impacto humano y tecnológico que supone, consideramos prioritario construir marcos teóricos flexibles para analizar su estatuto, su funcionamiento y los efectos concretos de la incorporación generalizada de estos modelos de lenguaje a nuestra existencia individual y colectiva. Dado el carácter exploratorio del trabajo, no pretendemos ofrecer respuestas definitivas ni unívocas, sino la apertura de un espacio de reflexión y debate en torno a estas máquinas conversacionales.

1. Funcionamiento de ChatGPT¹

¿Qué es ChatGPT?

Soy un modelo de lenguaje desarrollado por OpenAI llamado ChatGPT, diseñado para comprender y generar texto en base a las solicitudes de los usuarios. En esencia, soy un programa de inteligencia artificial entrenado para responder preguntas, generar texto y proporcionar asistencia en una variedad de temas (OpenAI, 2022a).

Como vemos en su autodefinición y en otras explicaciones, se utilizan verbos como “comprender”, “entender”, “leer”, los cuales se refieren a acciones propias de la inteligencia humana y requieren de una serie de procesos neuronales específicos y muy complejos. Sin embargo, si bien ChatGPT usa el lenguaje natural de manera exitosa, lo único que hace es calcular las palabras más probables de una secuencia; por tanto, no conoce su significado ni comprende lo que representan, y tampoco tiene intención de comunicar.

Este modelo de lenguaje está basado en la tecnología de redes neuronales y aprendizaje automático, cuyo objetivo es aprender a realizar ciertas tareas. En el caso de ChatGPT, responde preguntas, genera texto y proporciona asistencia (según sus propias palabras). El primer paso de este procesamiento consiste en el proceso de tokenización. Para ingresar en el sistema, los textos son traducidos a números y la red neuronal los procesa para reconocer la mayor cantidad de patrones posibles. Así, cada vez que aparezca un determinado patrón, la máquina identificará un *token* (palabra o subpalabra) que catalogará para reconocerlo en cada aparición. Sin embargo, estas etiquetas numéricas por sí solas no logran que el modelo incorpore aspectos semánticos. Para ello, es necesario el proceso de *embedding*, por el cual cada *token* se representa como un vector numérico de longitud fija, capturando, de esta forma, el significado semántico y contextual de cada uno en relación con las demás unidades del sistema. Esto es similar a una nube de palabras, en la cual la ubicación de cada elemento se encuentra determinada por su significado.

En GPT-3, la ubicación de cada *token* en el espacio está representada por un vector de 300 dimensiones, al modo de marcadores semánticos. Por ejemplo, los *tokens* “mujer”, “chica” y “gata” están marcados como “femenino”; “mujer” y “chica” también están marcados como “humano”, a diferencia de “gata”. Por su parte, ChatGPT da un paso más en este proceso, pues toma secuencias de varios *tokens* (frases) y los organiza en función de más de mil

¹ En este apartado, utilizaremos las explicaciones presentadas en videos de canales de YouTube especializados en tecnologías digitales: Dot CSV (2020a, 2020b, 2023) y Nate Gentile (2023). Si bien se han sucedido nuevas versiones de ChatGPT, en todas ellas se utiliza el procesamiento de lenguaje descrito en este artículo, pues las nuevas versiones mejoran en la arquitectura, la capacidad de procesamiento y la calidad de las respuestas, pero siempre utilizando los mismos procesos.

dimensiones. No obstante, no guarda frases completas sino simplificadas. A través de un proceso de limpieza del texto, elimina elementos vacíos o pobres en términos semánticos (artículos, verbos copulativos, preposiciones, mayúsculas, puntuación, etc.). Además, lleva a cabo el proceso de lematización, reemplazando las palabras flexionadas o derivadas por sus lemas (por ejemplo, una forma verbal conjugada por su infinitivo). El resultado es un *embedding* compuesto de secuencias simplificadas, comprimidas, en vinculación con otras de sentido similar.

De este modo, cuando hacemos una pregunta o damos una instrucción, el modelo la analiza, la convierte a su sistema de *embedding* y define su respuesta en función de una relación semántica. Como las secuencias guardadas no son gramaticalmente aceptables para poder responder, debe reponer palabras vacías y formas flexionadas, siguiendo una serie de reglas sintácticas aprendidas en su fase de entrenamiento. Por eso, el resultado es ligeramente distinto a los textos de entrada. Este proceso se lleva a cabo en la fase de *sampling*, por la cual el modelo utiliza sus *embeddings* para generar una distribución de probabilidad sobre el vocabulario y predecir cuál podría ser la siguiente palabra en una secuencia. Esto implica que la nueva frase estará lo suficientemente cerca de la original para que sea coherente con las intervenciones del usuario.

Por otra parte, el ChatGPT, a diferencia de GPT-3, incorpora los *transformers*, una tecnología creada por ingenieros de Google, que proporciona una nueva forma de organizar las redes neuronales, añadiendo una capa de atención. Ésta permite analizar una gran cantidad de datos, prestando atención sólo en los detalles primordiales, con el fin de determinar el contexto de las frases, atender al significado de las palabras, extraer conceptos clave, determinar qué términos tienen más peso, hacer un resumen, responder por tema, y todo de forma simultánea (no secuencialmente como en los modelos anteriores). Toda esta información es tenida en cuenta y se va enriqueciendo a lo largo del intercambio, algo que antes era imposible, dado que los sistemas no contaban con procesos de memoria suficientes para procesar toda una conversación. Así, el entrenamiento se vuelve mucho más rápido y eficaz.

A su vez, los ingenieros de OpenAI se propusieron construir la red neuronal más grande de la historia. Cada red neuronal tiene parámetros y una serie de ajustes que se configuran para que el modelo funcione correctamente. Durante la fase de aprendizaje, el sistema mismo realiza esos ajustes para que los resultados de salida sean los esperados; por tanto, cuantos más parámetros se le indiquen, se vuelve más ajustable, complejo y flexible. Si para entrenar a GPT-2 se habían usado 1.5 billones de parámetros, para GPT-3 se utilizaron 175 billones. El corpus de entrenamiento contó con el repositorio de Common Crawl, el Web Text Background, Wikipedia y bibliotecas, entre otras fuentes. Aunque para GPT-2 y GPT-3, el aprendizaje había sido no supervisado (los datos ingresaban sin etiquetar y el modelo llevaba a cabo la tarea), ChatGPT constituyó una versión refinada, *fine-tuned*, gracias a una capa

extra de entrenamiento con aprendizaje supervisado. Los datos de entrada fueron etiquetados por humanos, quienes generaron respuestas de alta calidad, indicando qué era cada contenido.

De este modo, el sistema no se centró tanto en clasificar como en establecer relaciones entre los datos, generando información de altísima calidad. Por último, añadieron también la etapa de *reinforcement learning from human feedback* (aprendizaje reforzado de *feedback* humano), durante la cual ChatGPT elaboraba cuatro respuestas posibles para cada pregunta y un grupo de humanos expertos debía puntuarlas para que el modelo aprendiera los caminos más satisfactorios para el ser humano. Esta utilización de “parches humanos” resolvió el problema del control de aprendizaje de estos modelos. Sin embargo, esta solución no es escalable exponencialmente, lo que constituye una de las dificultades más grandes que deben superar los desarrolladores de IA generativas.

2. Dilemas semióticos

Si observamos el funcionamiento concreto de ChatGPT, podemos decir que es un sistema que produce discurso: realizaciones concretas de lenguaje natural, escritas u orales, que se emiten a través de una interfaz. Estos fragmentos de materialidad sensible (cuya emisión, no hay que olvidar, depende de una infraestructura digital y un código binario que no vemos) constituyen enunciados en los cuales podemos reconocer, según indica Filinich (1998), dos niveles: el del enunciado, aquello que es objeto del discurso, y el de la enunciación, “presupuesta por todo enunciado en la medida en que todo discurso proviene de un *yo* que destina su alocución a un *tú*” (p. 19).

Nos centraremos, en primer lugar, en el tipo de enunciación que presentan las respuestas de ChatGPT. A diferencia de las emisiones lingüísticas de otros sistemas, este modelo utiliza los pronombres personales *yo/tú* para establecer una interacción con el usuario. Si nos situamos en el ámbito del análisis del discurso, esta atribución de pronombres nos permite, en principio, reconocer un sujeto que enuncia y un destinatario. Como afirma Benveniste (2004a), “es en y por el lenguaje como el hombre se constituye *sujeto*” (p. 180). Por tanto, el ejercicio de la lengua es el fundamento de la subjetividad: “*yo* se refiere al acto de discurso individual en que es pronunciado, y cuyo locutor designa [...] La realidad a la que remite es la realidad del discurso” (pp. 182-183). Esto conlleva, por un lado, no confundir el sujeto de la enunciación con el autor empírico, y, por otro lado, entender que el sujeto hablante debe dirigirse necesariamente a otro, porque decir *yo* implica siempre la presencia de un *tú* (p. 181). Esta constitución discursiva del sujeto se observa tanto en el uso de pronombres personales como en los indicadores de deixis “que organizan las relaciones espaciales y temporales en torno al ‘sujeto’ tomado como punto de referencia” (p. 183). Por tanto, el

sujeto que dice *yo* supone la “instauración de un lugar desde el cual se habla, de un centro de referencia alrededor del cual se organiza el discurso” (Filinich, 1998, p. 15).

Como dijimos, ChatGPT utiliza la dupla *yo/tú* en sus respuestas, por lo que estas categorías de la lingüística, pensadas siempre en función del discurso humano, se resignifican y complejizan. Por ejemplo, si le preguntamos: “¿Dónde estás?”, responderá: “Estoy aquí para ayudarte desde donde estés tú. ¿En qué puedo asistirte hoy?” (OpenAI, 2022b). Este ejemplo nos permite observar el uso de diversas clases de déicticos (pronombres, verbos conjugados, adverbios, etc.) que establecen claramente las dos posiciones del discurso y las coordenadas temporo-espaciales de la enunciación, aunque con sus peculiaridades. “Aquí” remite a la plataforma (interfaz) desde la cual acceder al chat, y, a la vez, a cualquier espacio donde el usuario se encuentre, lo que da cuenta de una especie de omnipresencia: “desde donde estés tú”.

Por otra parte, nos indica el presente, “hoy”, como punto de referencia constante de su discurso y, si bien en cualquier discurso el tiempo propio de la enunciación es el presente, ChatGPT pareciera atrapado en un presente eterno, dado que el modelo no experimenta verdaderamente ni el tiempo ni el espacio, ni posee sentido histórico; sólo cuenta con una memoria de información que abarca la totalidad de la conversación actual y el historial de conversaciones anteriores. Prueba de esto es que si le preguntamos: “¿Dónde estabas?”, responderá casi lo mismo que en la pregunta anterior, modificando sólo los tiempos verbales: “Estaba aquí, disponible para ayudarte en lo que necesitaras. ¿En qué puedo asistirte hoy?” (OpenAI, 2022b). Por tanto, pareciera que ChatGPT maneja coordenadas temporo-espaciales, en algún sentido, absolutas, pues su “aquí” y “hoy” no son susceptibles de matizaciones. Es el usuario quien, con sus propias intervenciones, complejiza estos aspectos, debido a que es la causa inicial de cualquier enunciación maquínica, a través de *prompts* de lenguaje natural.²

En este punto, conviene explicitar lo obvio: este sujeto que dice *yo* no es un ser humano, sino un modelo algorítmico. Esta particularidad con la cual la teoría lingüística no se había enfrentado hasta el momento —pues estudiaba la comunicación entre humanos—, nos plantea una serie de desafíos sobre la enunciación en la era digital. Proponemos hablar, entonces, de “enunciación maquínica” para señalar expresamente la distinción respecto de la humana, dada la alteridad radical de al menos uno de los sujetos que interactúan. En este sentido, las respuestas elaboradas por el chat constituirán un tipo discursivo nuevo, definido por su naturaleza digital, lo que implica una instancia de programación y diseño, basada en un código binario y una infraestructura material que la sustenta. Además, requiere una instrucción brindada por un usuario para generar texto y lograr su mejora y actualización

² “*Prompt*” es la instrucción que se proporciona a una herramienta de IA generativa para guiar su generación de respuestas o resultados.

permanente. Sus respuestas son el resultado de operaciones matemáticas que mencionamos en el punto 1, las cuales generan frases relacionadas semánticamente entre sí; por eso, en ocasiones, las respuestas son confusas o incluso circulares.

Aquí nos limitamos a indicar únicamente algunas de estas características formales, por una cuestión de extensión. Este tipo discursivo posee, además, condiciones propias de circulación e interpretación. Conocer la “autoría” de ChatGPT dará lugar a una recepción diferente respecto del discurso humano, pues no sólo está ligada a los rasgos indicados, sino también a las decisiones que los usuarios toman respecto de dichas interacciones: cómo se llevaron a cabo, cómo se interpretaron, cómo y por qué se pusieron en circulación, si son verdaderas, con qué objetivo fueron generadas, etc. Por ejemplo, en nuestra experiencia docente, los fragmentos copiados y pegados de ChatGPT que algunos alumnos insertan en sus trabajos son fácilmente reconocidos por un ojo avezado, y generan una serie de efectos en la evaluación. Pero ¿qué sucederá cuando no seamos capaces de distinguir el origen de ese discurso? En nuestra cultura de fuerte impronta digital, este tipo de discursos circula cada vez con mayor frecuencia; su reconocimiento como productos de una fuente de enunciación maquínica se vuelve un dato ineludible para interpretarlos de forma adecuada.

Antes de avanzar, nos referiremos brevemente a la producción de sentido. ChatGPT no entiende lo que dice, ni apela a generar sentido; sólo realiza operaciones matemáticas a partir de números. Ahora bien, en su procesamiento, hay un tratamiento semántico del lenguaje, tanto en lo relativo a los procesos de *embedding*, *sampling* y *transformers*, como en el corpus de entrenamiento. En cada intercambio concreto, por tanto, la reconstrucción del sentido corre por cuenta del usuario, porque es el único de los dos actores que tiene esta capacidad, aunque esa tarea está fuertemente orientada tanto por las operaciones semánticas que realiza el modelo como por su capacidad de recordar el desarrollo de la conversación y utilizarla como contexto. La situación comunicativa se entabla, de este modo, gracias al *prompt* lingüístico del usuario, con base en lo cual el modelo responde; de modo que es en el usuario en quien recae la tarea de obtener sentido del intercambio.

En cuanto al enunciado maquínico, suele estar lingüísticamente bien construido; es uno de los logros de ChatGPT. Para destacar algunas peculiaridades, apelamos a la teoría de Bajtín (2002) sobre los cuatro rasgos distintivos del enunciado, ya que nos permite reflexionar más allá de los términos puramente lingüísticos. En principio, las respuestas del modelo presentan los dos primeros rasgos, alternancia y conclusividad,³ sin mayor dificultad. Sin embargo, el

³ Bajtín (2002) define las fronteras del enunciado según un criterio de alternancia de sujetos. El enunciado termina cuando hay un cambio de sujeto en la interacción (p. 260), algo claramente delimitable en los intercambios con ChatGPT. Por otro lado, la conclusividad alude al hecho de que el hablante dijo todo lo que quería decir (p. 265), lo cual también se puede observar en las intervenciones del Chat. De hecho, utiliza fórmulas lingüísticas concretas de inicio y cierre en su intervención.

problema se plantea al considerar el enunciado como una postura activa del hablante dentro de una esfera de objetos y en relación con lo dicho anteriormente sobre ellos (p. 274). ChatGPT no tiene la capacidad de posicionarse ideológicamente, pero en sus respuestas se puede observar este posicionamiento que deriva del enorme corpus textual con el que fue entrenado y de las intervenciones de los “parches” humanos en los procesos de *fine-tuned* y *reinforcement learning from human feedback*. En ambos casos, intervienen personas que elaboraron los textos o realizaron acciones correctivas desde posiciones valorativas definidas y con una visión de mundo que impacta de lleno en el modelo: “Cada conjunto de datos [dataset] usado para entrenar sistemas de aprendizaje automático, tanto supervisado como no supervisado, tanto técnicamente sesgado como no, contiene una visión de mundo” (Crawford, citado en Carrión *et al.*, 2023, p. 25). En este sentido, podríamos pensar en el fenómeno de “polifonía” teorizado por Bajtín (1993) dado que, detrás de esas respuestas, se puede rastrear una gran cantidad de voces, situadas en coordenadas temporo-espaciales, culturales e ideológicas definidas, que se entretrejen en el discurso generado por el modelo. Esta máquina conversacional, por tanto, no habla de forma neutra, sino que constituye una voz-otra que se incorpora al gran diálogo universal de la especie humana. Y esto es más certero aún si pensamos con Carrión que estas máquinas, que aspiran a escribir como nosotros, no son más que máscaras que creamos y usamos (2023, p. 47), gracias a un proceso complejo de mediación tecnológica.

En relación con la postura activa del hablante respecto de su enunciado, Bajtín (2015) abre en sus escritos una línea de reflexión ética, que nos parece interesante traer a colación: la responsabilidad que el hablante posee respecto de su discurso. En el caso de ChatGPT, tenemos una entidad que no puede hacerse responsable de sus actos, pues no tiene conciencia ni autonomía ni agentividad. ¿Quién/es, entonces, se hace/n responsable/s de los dichos discriminadores de una IA generativa? En el caso de Tay, el bot adolescente de Microsoft que se volvió racista y xenófoba antes de aprender a conversar correctamente, fue la empresa la que debió hacerse cargo y retirar el chat de la web, asumiendo las consecuencias, no sin antes culpar a los usuarios de haber realizado un “esfuerzo coordinado [...] para abusar de las capacidades de conversación de Tay para que respondiera de forma inapropiada” (BBC News, 2016). Otro caso aún más delicado es el de un hombre belga que, en marzo del año pasado, se suicidó luego de un intercambio frenético de varias semanas con Eliza, un chatbot basado en GPT-J. La insuficiente respuesta del fundador de la plataforma cuestionada, Joseph Weizenbaum, fue “incluir un aviso dirigido a las personas que tengan pensamientos suicidas” (Página 12, 2023). Como afirmó el Ministro de Digitalización belga, Mathieu Michel, “en el futuro inmediato es fundamental identificar claramente la naturaleza de las responsabilidades que pueden haber llevado a este tipo de hechos” (Página 12, 2023). En todos los casos, queda claro que estos modelos están diseñados, programados y usados por seres humanos, con su propia visión de mundo y sus sesgos. Por tanto, la responsabilidad que atañe a la enunciación de discurso y a sus efectos no puede encontrarse en la máquina, la cual no posee la capacidad de responder con base en criterios

éticos. Esto es algo que los desarrolladores no han logrado resolver de forma definitiva (Chomsky, citado en Alonso, 2023).

Ahora bien, es necesario reparar en la última capa de entrenamiento permanente con la que cuenta ChatGPT: los intercambios con el propio usuario, pues sus respuestas buscan adaptarse, en el mayor grado posible, a lo que éste desea escuchar. Esta característica ha alcanzado una mejora significativa con la nueva versión GPT-4 Turbo, que permite crear versiones personalizadas de ChatGPT, entrenándolas a través de la conversación. Esta nueva etapa se relaciona con el cuarto rasgo del enunciado propuesto por Bajtín: la propiedad de estar destinado: cuando el hablante construye y emite un enunciado lo hace teniendo en cuenta las posibles reacciones de respuesta (2022, p. 285). Esta propiedad se da en ChatGPT de una forma absoluta. El modelo está programado para “capturar siempre nuevos datos que los mismos usuarios con su simple interacción producen, ya sea aceptando lo que la plataforma les da como respuesta o no aceptando y poniendo comentarios que la contraríen” (González Montaña, 2023, p. 42), con el único objetivo de satisfacer las expectativas del usuario y adecuarse a sus modos de pensar. Esto supone una gran diferencia respecto de la comunicación entre humanos y un problema para el desarrollo de los vínculos. El aumento de la frecuencia de diálogo con IA generativas customizadas nos distanciará cada vez más de nuestros pares, que no responden aquello que quiero escuchar, sino lo que quieren decir, propiciando una negociación de sentidos, propia de la comunicación humana y ampliamente saludable en términos de convivencia social. En el caso que analizamos, el usuario pregunta y el chat responde, a partir, entre otras cosas, de los intercambios con el propio usuario. Si el humano toma las respuestas como válidas, el proceso recursivo producirá una “relación cerrada”, pues ya no se podrá distinguir entre lo que piensa el usuario y lo que dice ChatGPT, hasta llegar al extremo de que aquél replicará las premisas de éste pensando que lo que reproduce es su propio criterio (González Montaña, 2023, p. 45).

Por último, en relación con la recepción, nos interesa retomar la instancia de interpretación de las respuestas del chat. El tono cordial y cercano (en ocasiones, divertido), el grado de atención que parece mostrar y la rapidez y pertinencia que poseen sus respuestas, pueden generar confusiones en los usuarios. Como hemos dicho, el chat no es un interlocutor humano y su capacidad lingüística es sumamente limitada, resultado de parametrizaciones que poco tienen que ver con el ejercicio del lenguaje natural. En la simulada acción de comunicar, no hace más que traducir en términos amigables una compleja trama de operaciones algorítmicas y *big data*, que nos abrumaría de sólo considerar el número de ítems que lo constituyen y la cantidad de servidores que se requieren para su procesamiento.

Pero ¿qué sucede si el usuario distraído o ingenuo aborda la conversación, olvidando que se comunica con una máquina? En principio, puede considerar esas respuestas como verdaderas y confiables; puede sentirse tentado de delegarle su propia capacidad de buscar información, construir respuestas, analizar críticamente la realidad, etc., quizá por cierta pereza o por falta

de autoestima al considerarla más eficiente y menos falible en esta tarea (Sadin, 2020; 2023a; 2023b). Por ello, es necesario formarse y generar conciencia, al menos, sobre dos cuestiones. La primera de ellas ya la mencionamos: ChatGPT no tiene parámetros de verdad ni éticos, sólo agrupa frases relacionadas semánticamente; por tanto, no puede ser considerada una instancia de autoridad en términos de conocimiento, ni una especie de gurú para la toma de decisiones. Es simplemente una herramienta. La segunda cuestión se refiere al hecho de que el modelo, si bien predice, no tiene una verdadera proyección de futuro, pues sólo considera elementos pasados con base en los cuales reconoce patrones y elabora nuevos textos siguiendo las mismas lógicas: “Hablamos de un patrón intuido o comprobado en un extraordinario conjunto de datos. Por eso es probable que acierte con sus respuestas” (Carrión *et al.*, 2023, p. 26). Su capacidad de informar, recomendar, predecir, deriva de este procesamiento incesante, que se va perfeccionando, pero que no tiene (al menos por ahora) la posibilidad de combinar elementos de una manera totalmente nueva para generar un salto cualitativo que modifique el rumbo de una disciplina, de un campo social o de un movimiento artístico.

Como hemos observado en otro trabajo (Lucifora, 2022), el riesgo no es sólo que los sistemas que utilizamos no puedan producir estos saltos de sentido, sino que, en la interacción con ellos, nuestra potencialidad para generarlos se vea reducida, porque nos obligan a interactuar siguiendo sus marcos sumamente esquematizados de procesamiento de lo real, estrechando los límites de nuestro pensamiento y de nuestra acción, tanto a nivel individual como cultural.

Conclusiones

El riesgo de estos avances tecnológicos no es el futuro distópico de nuestra aniquilación, sino los efectos individuales y culturales que ya poseen: los cambios que producen en nuestro modo de pensar(nos), vincularnos y actuar como seres humanos. En este sentido, consideramos necesario diferenciar claramente la enunciación humana de la enunciación maquínica, dado que ambas poseen un estatuto diverso y requieren de estrategias de interpretación distintas. Esta distinción nos resguarda de caer en la trampa de analizar los fenómenos del mundo físico y los objetos tecnológicos que los imitan con los mismos esquemas cognitivos y semióticos, y desde una perspectiva acrítica. Esta confusión es promovida por la industria *tech*, que utiliza, con objetivos principalmente comerciales, la táctica de asimilar sus productos de base binaria y algorítmica a nuestras experiencias cotidianas para facilitar la apropiación que hacemos de ellos. Estas coincidencias terminológicas producen fuertes reordenamientos de sentido (y, por tanto, de comportamiento) en la cultura, lo que genera confusión: “La IA no es un cerebro eléctrico [...] El órgano del pensamiento es un enigma. El problema es llamarla inteligencia —la cual radica en el cerebro— porque humaniza a la máquina” (Varsavsky, 2023).

En consecuencia, si hablamos de enunciación maquínica, es necesario caracterizar y reconocer un nuevo tipo de discurso, dada su circulación creciente en nuestra cultura, cuyos efectos causarán cambios definitivos. Si bien son máquinas que no producen significados de por sí, llevan a cabo un tratamiento semántico del lenguaje a través de operaciones matemáticas que orientan el sentido que los usuarios obtienen del intercambio. Por otra parte, puesto que un sistema informático no tiene la capacidad de tomar una posición ideológica, igualmente posee un impacto en ese sentido, pues los textos con los que fue entrenado, así como las personas que colaboran en su aprendizaje, poseen una visión de mundo heredada por el sistema. Queda abierta la posibilidad de analizar qué sucede cuando esas visiones entran en conflicto. Otra instancia de posicionamiento es el aprendizaje que el modelo lleva a cabo a partir de los intercambios con el usuario, buscando adaptarse lo mejor posible a sus expectativas. En este punto, consideramos prioritario estudiar los efectos que esta característica producirá en las relaciones interpersonales, mucho más complejas y menos complacientes que las del chat.

Finalmente, advertimos la necesidad de formarse y generar conciencia respecto del modo de interpretación de las respuestas de ChatGPT, para evitar caer en una visión ingenua y acrítica que termine reduciendo el ejercicio de la creatividad humana. Si, como se nos sugiere, terminamos usándolo para la mayoría de los ámbitos de nuestra vida, nos convertiremos en seres que sólo dan instrucciones, a través de un lenguaje que se volverá puramente instrumental y que nos alejará, como afirma Sadin, “de hacer, de crear, de la política, de la lucha” (2023b). Está claro que la idea no es rechazar el uso de ChatGPT, sino tratar de entender su funcionamiento y darle el lugar de lo que es: una herramienta que debe ser utilizada con responsabilidad y a la que no debemos delegar las capacidades que nos distinguen como seres humanos, entre ellas, la facultad del lenguaje que, “surgida de nuestro espíritu y de nuestra sensibilidad, [nos hace capaces] de establecer vínculos activos con los demás y con la realidad” (Sadin, 2023a).

Referencias

- Alonso, L. (2023, 24 de abril). La crítica de Noam Chomsky al sistema de Inteligencia Artificial Chat GPT. *Cultura inquieta*. <https://culturainquieta.com/pensamiento/la-critica-de-noam-chomsky-al-sistema-de-inteligencia-artificial-chat-gpt-3/>
- Bajtín, M. (1993). *Problemas de la poética de Dostoievski*. México: FCE.
- Bajtín, M. (2002). El problema de los géneros discursivos. En *Estética de la creación verbal* (pp. 248-293). Buenos Aires: Siglo XXI.
- Bajtín, M. (2015). *Yo también soy: fragmentos del otro*. Buenos Aires: Godot.
- Benveniste, E. (2004a). *Problemas de lingüística general I*. México: Siglo XXI.

- BBC News (2016, 25 de marzo). Tay, la robot racista y xenófoba de Microsoft. https://www.bbc.com/mundo/noticias/2016/03/160325_tecnologia_microsoft_tay_bot_adolescente_inteligencia_artificial_racista_xenofoba_lb
- Carrión, J., Taller Estampa y Chat GPT-2 y 3 (2023). *Los campos electromagnéticos. Teorías y prácticas de la escritura artificial*. Buenos Aires: Caja Negra.
- Dot CSV (2020a, 30 de junio). *Intro al Natural Language Processing (NLP) #1 - ¡De palabras a vectores!* [Video]. YouTube. https://www.youtube.com/watch?v=Tg1MjMIVArc&list=PL2YILQR_G2CsKUG2wm2hoA4SUTV2a7zdP&index=1
- Dot CSV (2020b, 15 de julio). *Intro al Natural Language Processing (NLP) #2 - ¿Qué es un embedding?* [Video]. YouTube. https://www.youtube.com/watch?v=RkYuH_K7Fx4&list=PL2YILQR_G2CsKUG2wm2hoA4SUTV2a7zdP&index=2
- Dot CSV (2023, 13 de noviembre). *¡Nuevos Custom GPTs: El Multiverso de ChatGPT!* [Video]. YouTube. <https://www.youtube.com/watch?v=hLYw06LYWIU>
- Filinich, M. I. (1998). *Enunciación*. Buenos Aires: Eudeba (Col. Enciclopedia Semiológica).
- Gentile, N. (2023, 7 de marzo). *¿Cómo funciona ChatGPT? La revolución de la Inteligencia Artificial* [video]. YouTube. https://www.youtube.com/watch?v=FdZ8LKiJBhQ&list=PL2YILQR_G2CsKUG2wm2hoA4SUTV2a7zdP&index=3
- González Montaña, A. (2023). El dispositivo digital en la era del capitalismo algorítmico: el caso de ChatGPT. *Sintaxis*, (11), 33-48. <https://doi.org/10.36105/stx.2023n11.04>
- Lucifora, M. C. (2022). Tendencias de la semiosis maquínica en la semiosfera global contemporánea. *In Itinere*, 12(1), 47-62. <http://revistas.ufasta.edu.ar/index.php/initinere/article/view/238>
- OpenAI (2022a). ChatGPT (consultado el 25 de abril de 2024) [Modelo de lenguaje GPT-3.5] <https://chat.openai.com/>
- OpenAI (2022b). ChatGPT (consultado el 25 de abril de 2024) [Modelo de lenguaje GPT-3.5] <https://chat.openai.com/>
- Página 12 (2023, 2 de abril). Consternación en Bélgica por el suicidio de un hombre tras hablar con un chatbot de IA. <https://www.pagina12.com.ar/536810-consternacion-en-belgica-por-el-suicidio-de-un-hombre-tras-h>
- Sadin, E. (2020). *La inteligencia artificial o el desafío del siglo. Anatomía de un antihumanismo radical*. Buenos Aires: Caja Negra.

Sadin, E. (2023a, 7 de enero). Opinión: El robot conversacional ChatGPT y nuestra relación con el lenguaje. *Página 12*. <https://www.pagina12.com.ar/513824-el-robot-conversacional-chatgpt-y-nuestra-relacion-con-el-le>

Sadin, E. (2023b, 3 de mayo). ChatGPT es un movimiento civilizatorio de profundas consecuencias antropológicas. *Página 12*. <https://www.pagina12.com.ar/545281-eric-sadin-chatgpt-es-un-movimiento-civilizatorio-de-profund>

Varsavsky, J. (2023, 22 de julio). El malentendido apocalíptico con ChatGPT. *Página 12*. <https://www.pagina12.com.ar/569859-el-malentendido-apocaliptico-con-chatgpt>

Acerca de la autora

María Clara Lucifora es profesora adjunta de Semiótica (UNMDP), investigadora del INHUS (CONICET-UNMDP) y cofundadora de la Fundación Ariadna. Investiga sobre el impacto de la revolución tecnológica en la semiosis humana, especialmente el generado por la IA. Publicaciones: “Cooperación semiótica entre seres humanos e inteligencia artificial”, en *Revista Internacional de Humanidades* (2021); “Tendencias de la semiosis maquina en la semiosfera global contemporánea”, en *In Itinere. Revista digital de Estudios Humanísticos* (2022); “El impacto de la inteligencia artificial en la semiosis humana desde la teoría de Peirce”, en *Actas de las IX Jornadas Peirce en Argentina* (2023).

Texto recibido: 14/05/2024; Revisado: 10/06/2024; Aceptado: 14/06/2024

Contenido publicado en acceso abierto bajo una licencia Creative Commons
Atribución-NoComercial-CompartirIgual 4.0 Internacional (CC BY-NC-SA 4.0)



Benemérita Universidad Autónoma de Puebla

Seminario de Estudios de la Significación

Casa de la Palma, 4 Sur 303, Primer Piso, Col. Centro, C.P. 72000, Puebla. Pue., México.

Tel. +52 222 2295502, semioticabuap@gmail.com

<https://topicosdelseminario.buap.mx/index.php/topsem>