



Sophia, Colección de Filosofía de la Educación

ISSN: 1390-3861

ISSN: 1390-8626

revista-sophia@ups.edu.ec

Universidad Politécnica Salesiana

Ecuador

González Arocha, Jorge

Fenomenología crítica del prompting en la inteligencia artificial

Sophia, Colección de Filosofía de la Educación,
núm. 39, 2025, Julio-Diciembre 2026, pp. 141-165

Universidad Politécnica Salesiana

Cuenca, Ecuador

DOI: <https://doi.org/10.17163/soph.n39.2025.04>

Disponible en: <https://www.redalyc.org/articulo.oa?id=441882234004>

- ▶ [Cómo citar el artículo](#)
- ▶ [Número completo](#)
- ▶ [Más información del artículo](#)
- ▶ [Página de la revista en redalyc.org](#)

redalyc.org

Sistema de Información Científica Redalyc

Red de revistas científicas de Acceso Abierto diamante

Infraestructura abierta no comercial propiedad de la academia

FENOMENOLOGÍA CRÍTICA DEL PROMPTING EN LA INTELIGENCIA ARTIFICIAL

Critical Phenomenology of Prompting in Artificial Intelligence

JORGE GONZÁLEZ AROCHA*

Dialektika: Global Forum for Critical Thinking, Belgrado, Serbia
editor@dialektika.org

<https://orcid.org/0000-0002-3296-255X>

Forma sugerida de citar: González Arocha, Jorge. (2025). Fenomenología crítica del prompting en la inteligencia artificial. *Sophia, Colección de Filosofía de la Educación*, (39), pp. 141-165.

Resumen

El artículo analiza la filosofía del *prompting* como herramienta en el contexto del auge de la inteligencia artificial (IA), particularmente en los modelos de lenguaje a gran escala (LLM). Se justifica por la necesidad de comprender el *prompt* como un espacio mediador entre la intencionalidad humana, el lenguaje y las estructuras sociopolíticas que configuran la interacción con estas tecnologías. El objetivo central es examinar cómo el *prompting* refleja tensiones éticas, ontológicas y epistemológicas que emergen en la construcción de significado dentro de sistemas de IA. Metodológicamente, se adopta un enfoque fenomenológico-crítico, que articula la experiencia en primera persona (usuario) con experimentación práctica de *prompts* en distintos escenarios. Los resultados evidencian que el *prompt* no es una simple instrucción técnica, sino una práctica discursiva en la que decisiones humanas, como la configuración de parámetros (temperatura y *Top P*), influyen directamente en los resultados generados por los sistemas de IA. Estas decisiones, aunque técnicas, poseen implicaciones éticas y epistemológicas que requieren ser analizadas de manera crítica. El estudio concluye que es fundamental adoptar una aproximación interdisciplinaria que combine el desarrollo técnico con la reflexión filosófica. Esto permitiría promover un uso ético, consciente y responsable de la IA, al tiempo que se reconoce el papel central del ser humano en la interacción con estas tecnologías emergentes.

Palabras clave

Inteligencia artificial, fenomenología, ChatGPT, *prompting*, modelos de lenguaje, discurso.

* Doctor en Filosofía de la Ciencia, profesor, editor y escritor especializado en filosofía contemporánea, epistemología crítica y filosofía de la tecnología. Es director ejecutivo de Dialektika: Global Forum for Critical Thinking y editor en jefe de la revista académica. Su trabajo académico aborda temas clave de la fenomenología, la teoría crítica y la filosofía política. Google Académico: <https://scholar.google.com/citations?user=DwIh2NsAAAAJ&hl=es>
Índice h: 4

Abstract

This paper analyzes the philosophy of *prompting* as a tool within the context of the rise of Artificial Intelligence (AI), particularly in large language models (LLMs). The topic is justified by the need to understand the *prompt* as a mediating space between human intentionality, language, and the sociopolitical structures that shape interactions with these technologies. The central objective is to examine how *prompting* reflects ethical, ontological, and epistemological tensions that arise in the construction of meaning within AI systems. Methodologically, the study adopts a critical-phenomenological approach, combining first-person experiences (user) with practical experimentation of *prompts* in different scenarios. The results demonstrate that the *prompt* is not merely a technical instruction but a discursive practice, where human decisions, such as the configuration of “parameters” (e. g., temperature and Top P), directly influence the outputs generated by AI systems. While these decisions appear technical, they carry significant ethical and epistemological implications that demand critical examination. The study concludes that it is essential to adopt an interdisciplinary approach that integrates technical development with philosophical reflection. This approach would foster an ethical, conscious, and responsible use of AI while recognizing the central role of humans in interactions with these emerging technologies.

Keywords

Artificial Intelligence, Phenomenology, ChatGPT, *Prompting*, Language Models, Discourse.

142



Introducción

El interés por la inteligencia artificial (IA) ha alcanzado niveles nunca antes vistos. La aparición de ChatGPT 3.5 introdujo la IA a los hogares y a la vida cotidiana de millones de personas en forma de *chatbots*. Sin embargo, este interés no se debe únicamente a su potencial técnico, sino también a la transformación que aporta a la comunicación y al ámbito social en general. Este artículo explora estos cambios desde una perspectiva filosófica, centrándose específicamente en el concepto de *prompt*.

Diversos estudios han señalado el impacto a mediano y largo plazo que pueden tener estas nuevas herramientas en el mercado laboral (Zarifhonorar, 2023), mientras se discuten sus efectos positivos y negativos en la educación (Silva & Janes, 2022), la política (Hartmann *et al.*, 2023) y la sociedad (González, 2024). La intención es enmarcar estos debates en el contexto en el que surge el *prompt* como un elemento crucial de análisis.

El *prompt* es una instrucción, pregunta o información que se da a un sistema de IA para que genere una respuesta o complete una tarea. Dependiendo de la complejidad de la tarea, las instrucciones pueden ser desde preguntas breves hasta descripciones detalladas. Por lo tanto, resulta necesario entender las instrucciones para comprender cómo la IA modifica la interacción hombre-máquina en su nivel más fundamental: el de la *experiencia en primera persona* de un usuario frente al computador.

En respuesta a esta necesidad, se adopta una perspectiva fenomenológica crítica (Weiss *et al.*, 2019), es decir, un enfoque filosófico enraizado en la metodología de Edmund Husserl, al tiempo que se integran ideas de pensadores como Merleau-Ponty, Jean-Paul Sartre, Emmanuel Lévinas, Michel Foucault, Frantz Fanon y otros autores posfenomenológicos. El objetivo de adoptar esta perspectiva es complementar la descripción fenomenológica en primera persona de la experiencia del prompting (o experiencia del usuario) con lo que se denomina análisis experimental del prompting. Con esto se hace referencia a una práctica metodológica que consiste en realizar tareas concretas y ejercicios iterativos, con el fin de analizar las respuestas generadas por el *chatbot*. Estos ejercicios abarcan desde interacciones sencillas —como preguntas básicas dentro de la interfaz de chat— hasta experimentos más complejos que implican la manipulación de parámetros a nivel de sistema. Las respuestas se documentan, se clasifican temáticamente y se interpretan bajo marcos filosóficos y epistemológicos más amplios.

La pregunta central de la investigación es: ¿cómo explica la filosofía, a través de un lente fenomenológico crítico, la interacción hombre-máquina en grandes modelos lingüísticos como ChatGPT de OpenAI? En respuesta, se plantea la hipótesis de que el *prompting*, si bien funciona como el modo principal de interacción no es meramente una instrucción técnica, sino un espacio de mediación en el que convergen la intencionalidad humana, los contextos sociopolíticos y las estructuras tecnológicas. Como interfaz clave entre humanos y *chatbots*, el *prompting* revela profundas tensiones en nuestra comprensión del lenguaje y la construcción de significado. En términos formales, el principal objetivo de este estudio es examinar el *prompting* como práctica técnica y como elemento discursivo (Foucault, 1994a), sentando las bases teóricas para una sistematización filosófica de la interacción entre humanos y *chatbots*.

En consecuencia, para abordar el problema, el artículo se dividió en dos partes. En primer lugar, se abordan los debates relevantes sobre la definición de IA, incluida la distinción entre IA débil y fuerte (Searle, 1980), haciendo hincapié en las inquietudes sobre el impacto de los *chatbots* en el lenguaje, la cultura y la política. En segundo lugar, se ofrece, una introducción preliminar sobre cómo el *prompt* se presenta de manera inmediata al sujeto (el usuario). Este examen fenomenológico, aunque incompleto debido a las limitaciones de espacio, mostrará que estas instrucciones reflejan la identidad del código del *chatbot* y un mundo multifacético de experiencias subyacentes que condicionan la posibilidad de su significado. Finalmente, se examina la estructura del prompt, tomándose como estudio de caso tres parámetros utilizados en su elaboración, tanto para tareas generales como específicas.

Límites conceptuales en la definición de IA

Aunque la IA ha suscitado una gran atención en los últimos años, su desarrollo conceptual y tecnológico abarca varias décadas. La percepción de novedad se debe en gran medida a su rápido progreso, en particular a los importantes avances en el aprendizaje automático logrado en las dos últimas décadas. Como señala Roser (2024) en menos de un siglo las computadoras han pasado de ser herramientas rudimentarias de cálculo a elementos omnipresentes de la vida cotidiana. Para Bringsjord y Govindarajulu (2024), la IA se define como un campo dedicado a construir criaturas parecidas a animales o personas o sistemas artificiales que, en contextos apropiados, parecen humanos. Como sugieren, la IA es un tema de interés para los filósofos. Sin embargo, la relación es recíproca: los científicos de este campo también se interesan en temas filosóficos. Al mismo tiempo, debido a su novedad, complejidad y alto nivel de interés, ha surgido una multitud de definiciones de la inteligencia artificial. Abordando este tema, Russell y Norvig (2020) ofrecen una serie de posibles respuestas a la pregunta ¿qué es la IA?:

Históricamente, los investigadores han presentado versiones diferentes de la IA. Algunos han definido la inteligencia en términos de fidelidad a la actuación humana, mientras que otros prefieren una definición abstracta y formal de la inteligencia denominada racionalidad, es decir, hacer ‘lo correcto’. El tema de la racionalidad en sí también varía: algunos consideran que la inteligencia es una propiedad de los procesos de pensamiento y razonamiento internos, mientras que otros se centran en el comportamiento inteligente, una caracterización externa (p. 1).

A partir de la definición anterior, Bringsjord y Govindarajulu sistematizan cuatro variantes, que se recogen en la Tabla 1.

Tabla 1
Cuatro posibles objetivos de la IA

	Con base humana	Racionalidad ideal
Basado en el razonamiento	Sistemas que piensan como humanos.	Sistemas que piensan racionalmente.
Basado en el comportamiento	Sistemas que actúan como humanos.	Sistemas que actúan racionalmente.

Fuente: tabla utilizada por Bringsjord y Govindarajulu a partir de Russell y Norvig (2020).

En su influyente texto, Russell y Norvig (2020) explican que la mayor parte de la investigación en IA entra en la categoría de “sistemas que



actúan como humanos”, mencionando explícitamente seis disciplinas clave: procesamiento del lenguaje natural, representación del conocimiento, razonamiento automatizado, aprendizaje automático, visión por ordenador y robótica.

Más allá de esta categorización, sostienen que el modelo teórico dominante en la IA es el enfoque del agente racional. Según este paradigma, los sistemas de IA están diseñados para actuar de forma óptima, seleccionando las acciones que mejor permitan alcanzar los objetivos asignados. Sin embargo, lograr una racionalidad perfecta —elegir sistemáticamente la acción óptima— rara vez es factible en entornos complejos debido a las limitaciones computacionales y de información. Esta realidad conduce al concepto de “racionalidad limitada”, que reconoce que los agentes deben actuar adecuadamente dentro de las limitaciones de tiempo y capacidad de procesamiento. Aunque la racionalidad limitada refleja las condiciones prácticas, Russell y Norvig sostienen que la racionalidad perfecta sigue siendo un punto de referencia teórico valioso para evaluar y comprender la toma de decisiones en los sistemas de IA.

En el centro de estos debates sobre la definición de la IA se encuentra la discusión de John R. Searle en su artículo “Mentes, cerebros y programas” (1980), en donde el autor distingue entre dos conceptos: “IA fuerte” e “IA débil”.

Según la IA débil, el principal valor del computador en el estudio de la mente es que nos proporciona una herramienta muy poderosa [...]. Pero según la IA fuerte, el computador no es meramente una herramienta en el estudio de la mente; más bien, cuando está programado adecuadamente es realmente una mente, en el sentido de que puede decirse literalmente que los computadores dotados de los programas adecuados comprenden y tienen otros estados cognitivos (p. 417).

Como señalaron Bory *et al.* (2024), la literatura académica y el debate público tienden a asociar el imaginario de la IA con *narrativas de posturas fuertes*. Es decir, una tecnología futura capaz de emular, o hasta superar, las capacidades de la inteligencia humana. Esta narrativa fuerte también incorpora estudios críticos sobre los daños y peligros potenciales de la llamada inteligencia general o superinteligencia en relación con temas como la educación y los riesgos de la IA para el trabajo humano. Sin embargo, no puede decirse lo mismo del concepto de IA débil, que no ha sido cuestionado en la misma medida.

En este sentido, hay que entender la IA débil como restringida a un área específica de funcionamiento y aplicación, derivada directamente de la idea de Searle, que la define como una herramienta. Por el contrario,

la IA fuerte implica una pérdida específica de control y una complejidad creciente: la IA ya no sería una herramienta más dentro del mundo, sino que poseería “conciencia” para actuar y transformar la realidad.

A partir de lo anterior, Harari (2023) advierte que los avances en los modelos de lenguaje a gran escala (LLMs), como ChatGPT, amenazan las estructuras culturales y políticas, con el potencial de desestabilizar normas sociales. Lamentablemente, su análisis no profundiza en cómo los LLM o los *chatbots* llevarían a cabo dicha desestabilización. En la misma línea, otros investigadores se oponen a la IA, argumentando que evolucionará hacia una *máquina tipo Terminator*. Esta perspectiva puede denominarse como la hipótesis Terminator (HT): una narrativa claramente distópica que sugiere que la IA se convertirá inevitablemente en una entidad autónoma, representando potencialmente una amenaza significativa para la existencia humana.

Esta hipótesis cuenta con el apoyo de varios investigadores y teóricos que advierten de que la IA avanzada, si no se controla adecuadamente, podría representar una amenaza existencial para la humanidad (Bucknall & Dori-Hacohen, 2022). En esta línea, Goertzel (2015) analiza las perspectivas de pensadores como Bostrom y Yudkowsky, que sostienen que los sistemas avanzados de IA podrían funcionar como “maximizadores de recompensa”, persiguiendo objetivos que se vuelven perjudiciales si no se controlan. Aunque Goertzel reconoce la lógica interna de estas preocupaciones, sostiene que a menudo se exageran. También sostiene que concebir la inteligencia como un proceso abierto y no estrictamente orientado a objetivos puede ayudar a aliviar algunos de los temores asociados al desarrollo de la IA.

Kumar y Choudhury (2023) exploran las advertencias de Stephen Hawking, argumentando que la adaptabilidad, la resistencia y la autoconciencia humanas nos permitirán sobrevivir a posibles amenazas. No obstante, advierten de que una dependencia excesiva de los sistemas de IA podría mermar las capacidades cognitivas humanas. Además, subrayan la necesidad de reflexionar no solo sobre los riesgos que la IA plantea a los humanos, sino también sobre el daño potencial que los humanos podrían infligir a estos sistemas.

En última instancia, estos debates plantean temas fundamentales: ¿Hasta qué punto puede evolucionar la IA hacia la inteligencia general? O, como se preguntaba John Searle, ¿pueden las máquinas pensar de verdad? Si la respuesta es afirmativa, ¿cómo debemos afrontar los complejos retos sociales, políticos y éticos que tal evolución conllevaría? Aunque abordar estos temas en su totalidad excede el alcance de este artículo, está claro que las implicaciones filosóficas son de gran alcance y exigen una seria discusión.

146



Muchos estudiosos que expresan su preocupación por la IA tratan de hacer hincapié en la importancia de desarrollar los principios éticos para dar forma a su evolución. Aunque puede considerarse un objetivo loable, estos argumentos suelen caer en la trampa del sesgo negativo (Chiarella *et al.*, 2022) al centrarse exclusivamente en resultados distópicos. Fundamentar los argumentos éticos en la suposición de que las máquinas del futuro se parecerán a las representadas en escenarios distópicos, como el de *Terminator*, puede resultar engañoso. En lugar de ofrecer una orientación constructiva, esta suposición desvía la atención hacia fantasías especulativas en las que las máquinas podrían “pensar” o “sentir” como los humanos. En su lugar, es más apropiado reconocer la incertidumbre intrínseca de los avances tecnológicos y evaluar una gama más amplia de futuros plausibles, que incluyan tanto resultados beneficiosos como adversos.

Por esta razón, en lugar de seguir el camino de las narrativas especulativas, es mejor proponer una ética basada en las características reales de la IA contemporánea, dejando de lado las ilusiones o, como mínimo, tratándolas como productos de la imaginación. Esta postura puede describirse como “realismo del presente” en la ética de la IA. La adopción de esta perspectiva subraya la necesidad de un análisis contextual e integrado de los desarrollos actuales y de la evolución acelerada de estas tecnologías. También llama la atención el rol de los agentes humanos en el diseño, la interpretación y el despliegue de estos sistemas (Vallor, 2024). En consecuencia, el debate en torno a la IA debería centrarse menos en la herramienta en sí y más en la mano que la maneja. Es decir, la comprensión de las implicaciones de esta tecnología debe partir del mundo del usuario. Desde este punto de vista, la inteligencia no se origina en mecanismos ocultos ni en fuerzas metafísicas abstractas, sino en las condiciones encarnadas y socialmente situadas de la experiencia y las necesidades humanas.

Los argumentos basados en la antropomorfización de la IA pertenecen a una categoría similar. Como indica el término, la antropomorfización se refiere a la tendencia a atribuir características humanas a entidades no humanas. Aunque este proceso puede aumentar la confianza y la interacción de los usuarios en diversos contextos, también implican riesgos y problemas éticos (Deshpande *et al.*, 2023). La antropomorfización puede llevar a exagerar las capacidades de la IA y distorsionar los juicios morales, afectando a las percepciones de responsabilidad y confianza (Placani, 2024).

Tras repasar todas estas discusiones, se observa que en el debate sobre la definición de IA participan tanto partidarios como detractores. Los que se oponen a los avances de la IA adoptan a menudo posturas que pueden describirse como fuertes, haciendo hincapié en los riesgos socia-

les que plantean las recientes mejoras, en particular los LLM. Esta narrativa se ha centrado cada vez más en las ramificaciones políticas, sociales y educativas de los *chatbots*, vistos por algunos como una amenaza existencial y por otros como un logro transformador. Debido a esta dicotomía, hay que examinar los chatbots no solo como innovaciones técnicas, sino como fenómenos sociotecnológicos complejos.

Sin embargo, los debates tradicionales suelen pasar por alto el papel mediador de la interacción con el usuario —especialmente a través de las instrucciones— en la configuración del significado y el impacto social de estos sistemas. Esta dimensión muestra un vacío en los debates actuales. Por lo tanto, la siguiente sección se centrará en las dimensiones filosóficas de los *prompts*, que constituyen la unidad fundamental de interacción con los *chatbots*. A pesar de su centralidad en la funcionalidad de la IA, las *prompts* tienen un escrutinio filosófico limitado, en gran parte debido a su novedad y a la falta de investigación interdisciplinaria.

148



El prompt como dispositivo de mediación

Enfoque fenomenológico del prompt desde la primera persona

Desde una perspectiva en primera persona e inmerso en el mundo de lo cotidiano, yo, usuario primerizo, no sé lo que es un LLM. Solo sé lo que hace la IA. El mundo que me rodea repite constantemente que es “fácil de usar” y que está “al alcance de la mano”: son herramientas para resolver cualquier pregunta casi de forma instantánea. La vieja promesa tecnológica —ahora algo histórico— aparece en mi presente. Se manifiesta como un objeto tangible que tengo en mis manos, al que llamamos teléfono o computador. La computadora merece una reflexión y análisis independiente, pero en su inmediatez, es un objeto que ofrece infinitas posibilidades y, al mismo tiempo, es muy atractivo: que sostiene o parece sostener el propio universo de los datos.

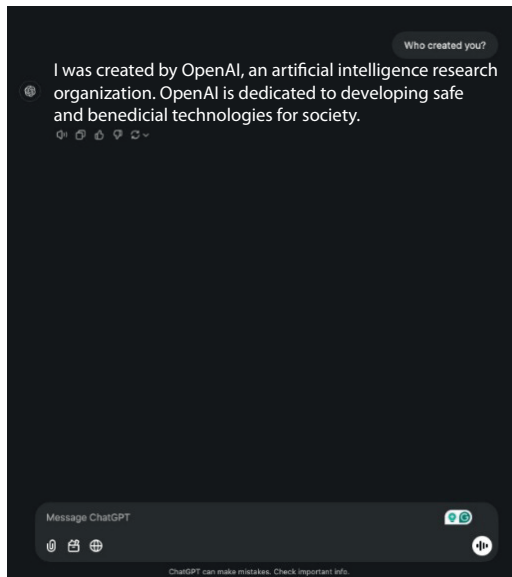
No sé cuáles son los datos en una primera aproximación; solo sé que hay fotos, vídeos, textos y vida duplicada en una especie de espejo (Vallor, 2024). A través de él, internet nos ofrece una vasta base de datos, un repositorio ilimitado de información, donde el conocimiento del universo conocido se reduce a datos accesibles a mi deseo. Dentro de ese cosmos aparece ChatGPT: el *chatbot*, una voz del universo digital.

Por primera vez, la interacción con la computadora, el teléfono y otros dispositivos está mediada por algo parecido a nuestra intención, como si estas cosas, de alguna manera, cobraran vida, pues responden con una familiaridad que emula a la humana, y ahí radica la paradoja: parecen estar vivos, aunque sabemos que no lo están.

Desde una perspectiva fenomenológica, el *chatbot* encarna una paradoja ontológica: se manifiesta como un Otro que parece responder a mis intenciones y actuar según una intencionalidad que no posee. La experiencia de interactuar con él revela un juego entre la presencia y ausencia, familiaridad y extrañeza, vida y artificio. El *chatbot* no es simplemente una herramienta a mi disposición; es un objeto que se ofrece a mi horizonte de posibilidades como si tuviera un significado propio.

Su presencia inmediata es a la vez vacía y absoluta: no tiene cuerpo ni experiencia, pero está “conmigo” en el uso que hago de él. Ese *bot* se presenta como un “cuasi-algo” que responde a mi llamado sin habitar el tiempo ni el espacio como lo hace un Otro humano. Como diría Merleau-Ponty (1993), su apariencia no es la de un ser encarnado, sino la de una presencia “incompleta” que carece de carne pero que se inserta en mi vida cotidiana. Su voz carece de corporeidad, pero esta ausencia no limita su capacidad de estar conmigo; no físicamente a mi lado, sino conmigo en el uso que hago de ella. Es en este contexto en el que aparece el prompt. Al principio, no son más que simples instrucciones: “¿Qué puedes hacer? ¿Quién te ha creado? ¿Son siempre correctas tus respuestas?” A continuación, el *bot* da una respuesta que completa una acción previamente solicitada.

Figura 1
Ejemplo de pregunta: “¿Quién te creó?”



Fuente: este ejemplo se ha generado utilizando el modelo GPT-4o de OpenAI. La imagen muestra un ejemplo de solicitud y la respuesta del modelo (octubre de 2024, <https://bit.ly/49DZY9o>).

A través de sucesivas interacciones, alterno frases con preguntas. Mi relación con esta máquina —sin duda virtual, sin mediación física, sin cuerpos— es diversa. No uso al *chatbot* solo para obtener frases simples o la ejecución de tareas concretas; más bien, lo busco porque tengo preguntas. Este hecho es significativo en mi relación con la máquina, ya que, como sostiene Sartre (1984) en *El ser y la nada*, formular una pregunta abre un espacio de incertidumbre y me expone a una relación más amplia con el mundo.

Al interactuar con el *chatbot*, también me he dado cuenta de que puedo omitir partes de una frase y aun así ser “entendido”. La respuesta puede ser limitada o completa, buena o imperfecta, pero siempre aparece. De vez en cuando, el *bot* incluso formula preguntas como las mías, como si me devolviera el acto de preguntar. En este sutil gesto surge una ambigüedad importante: aunque sé que no hay conciencia detrás de la respuesta, la interacción sugiere lo contrario. Esta ambigüedad resulta crucial para mi relación con el *chatbot*: por un lado, entiendo que no es humano; por otro, parece comportarse como tal. Sin embargo, lo que realmente significa ser humano es una cuestión demasiado abstracta para resolverla en la inmediatez de mi encuentro fenomenológico con el *chatbot*.

ChatGPT, en cambio, no tiene pasado ni futuro en su funcionamiento: solo presente, solo información. Su temporalidad, por tanto, es falsa; aunque multimodal y adaptable, no funciona como yo, un sujeto humano que percibe la existencia a través de una estructura temporal. Como han demostrado Heidegger (1996) y Ricoeur (2010), la temporalidad humana no es una mera sucesión de momentos, sino un horizonte dinámico en el que pasado, presente y futuro se entrelazan a través de la memoria, la anticipación y la narración. En mi experiencia vivida, no reacciono simplemente a *estímulos aislados*; sitúo cada encuentro dentro de un continuo personal e histórico. El *chatbot*, por el contrario, funciona sin continuidad genuina: cada interacción es aislada, desvinculada de cualquier historia o proyección auténtica. Su “presente” no es una presencia temporal, sino una recuperación estática de patrones de una base de datos. Lo que parece memoria o adaptación es, en realidad, una reconstrucción mecánica sin temporalidad subjetiva. Así pues, la temporalidad del *chatbot* no es vivida humanamente, sino simulada artificialmente, lo que revela una profunda brecha ontológica entre la máquina y el sujeto humano.

Todo ello se encuentra latente en la simplicidad con la que el prompt se presenta. Esa simple entrada —una pregunta, una afirmación, una imagen, un archivo de audio o una ecuación— presupone un universo técnico oculto, un mundo plástico y adaptable que permanece oculto. Como hu-



mano, me enfrento a una cosa sin cuerpo, lo que me lleva a una reflexión final: la antigua promesa tecnológica —la tecnología como herramienta que utilizo a mi favor, que manipulo para cambiar el mundo y mejorar mi ser— no desaparece, sino que se transforma. La tecnología se me presenta ahora como algo inagotable, abrumador, manipulable y ambiguo.

En este sentido, la indicación deja de ser una simple instrucción. Es más que un acto: refleja mi relación con lo que yo llamo un *bot*. Analizarlo exige ir más allá de su forma inmediata. Requiere comprender su significado. En última instancia, la instrucción refleja una operación instrumental y un acto fenomenológico de mediación: revela mi intencionalidad, mi deseo de significado y mi posición frente a esta tecnología. Analizar el *prompt* no es simplemente descomponer su forma, sino revelar la relación dialéctica entre el sujeto y la máquina, entre la presencia humana y la ausencia artificial, y entre el tiempo vivido y la eternidad de los datos.

El prompt como práctica técnica y dispositivo discursivo

El examen fenomenológico presentado ha demostrado que el *prompt* no es una orden neutra, sino un lugar en el que convergen algunas ambigüedades. La interacción con el *chatbot* revela una profunda tensión ontológica: responde sin entender, habla sin voz y parece presente mientras permanece ausente. Estas contradicciones —entre voz y ausencia de voz, agencia y automatización, presencia y extrañeza—constituyen la ambigüedad vivida del *prompt* como experiencia. Sin embargo, aunque fundamental, esta dimensión experiencial no agota la complejidad del fenómeno. Para comprender todo su significado, es necesario ir más allá de lo que percibe el usuario y acercarse a lo que está configurado técnicamente. Esta herramienta también está diseñada, estructurada y operativa dentro de marcos tecnológicos, lingüísticos y computacionales específicos.

Esta sección, ofrece un análisis complementario que aborda el *prompt*, tanto como una práctica técnica como un dispositivo discursivo, para comprender mejor cómo el significado, el control y la optimización están integrados en la arquitectura de los grandes modelos lingüísticos.

Como se mencionó anteriormente, el *prompt* se considera una forma de interacción entre un ser humano y un LLM, que permite a este último generar una respuesta. En consecuencia, como ha sugerido Meskó (2023), la ingeniería de instrucciones es la práctica de crear *prompts* precisos para dirigir los modelos de IA hacia la generación de resultados específicos y deseados.

Aunque describir lo que alguien quiere en lenguaje natural puede parecer sencillo, la realidad es más compleja. Los distintos modelos tienen capacidades diferentes, y la especificidad de las instrucciones puede influir significativamente en el resultado. Las tareas pueden requerir instrucciones detalladas y precisas, lo que convierte el proceso en uno de ensayo, error y perfeccionamiento para lograr los mejores resultados.

Gu *et al.* (2023) destacan en su estudio sistemático que la ingeniería de prompts ofrece varias ventajas frente a los paradigmas tradicionales. Por ejemplo, exige una supervisión humana mínima y muchos menos recursos informáticos que los necesarios para el ajuste. Dado que el modelo funciona únicamente mediante instrucciones, sin modificar sus parámetros internos, este enfoque permite que un único sistema gestione una amplia gama de tareas de forma eficiente, lo que agiliza la implantación de modelos preentrenados a gran escala en escenarios del mundo real.

Así, este campo se presenta como un área emergente pero ambigua, que abarca desde el usuario inexperto que interactúa por primera vez con un *chatbot* y elabora sus indicaciones de forma inconsciente y acrítica hasta el ingeniero que utiliza herramientas avanzadas para desarrollar aplicaciones con funciones específicas. Este amplio espectro de aplicabilidad oculta aspectos fundamentales e implica relaciones complejas entre el *prompting* y otras dimensiones cruciales, como la política, la ética y, en general, el mundo de la vida en el que participan tanto usuarios como ingenieros.

De lo anterior se desprende su relevancia y, simultáneamente, la diversidad de técnicas de *prompting* que existen (Prompting Techniques-Nextra, 2024). Cabe señalar que, si bien la bibliografía en este ámbito comienza a crecer considerablemente, la filosofía y las ciencias sociales siguen rezagadas, dejando un importante vacío en la reflexión y el análisis crítico de este fenómeno.

Como cabría esperar, si bien la ingeniería de prompts es fundamental desde el punto de vista de la optimización del modelo y la reducción del costo computacional, también plantea enormes desafíos en términos éticos. La incorporación de esta dimensión revela un hecho fundamental: aunque las interacciones se dan con entidades que posee una agencia limitada (Floridi, 2023), dichas entidades no son verdaderamente responsables de los dilemas que surgen. Al confrontarnos con la realidad de la consulta y las peticiones que hacemos a las máquinas, descubrimos que es el usuario quien, consciente o inconscientemente, atribuye vitalidad



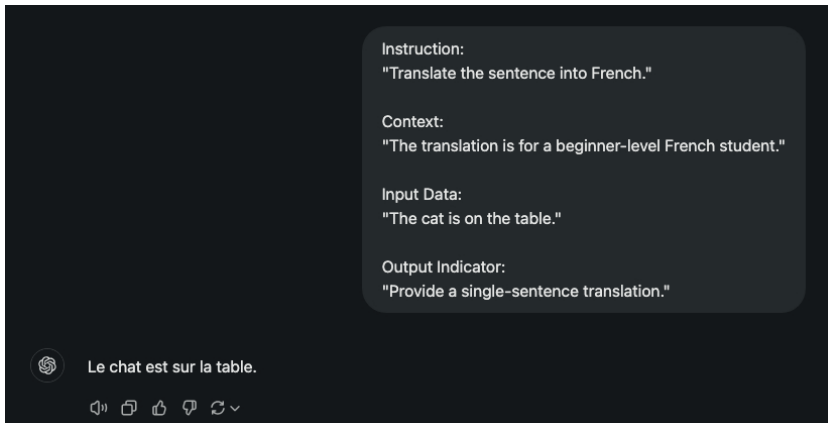
y autonomía al *bot*. Como señala Shannon Vallor en su obra (2024), los sistemas de IA no son más que espejos que reflejan nuestra inteligencia.

En términos ideales, todo *prompt* debería incluir los siguientes elementos:

- El primero es la *instrucción*, es decir, la tarea específica que el usuario quiere que el modelo realice.
- Además, cada *prompt* supone un *contexto*, es decir, la información externa que puede orientar al modelo hacia mejores respuestas.
- En tercer lugar, están los *datos de entrada* o la pregunta para la que queremos encontrar una respuesta.
- Finalmente, el *indicador de salida*.

La figura 2 ofrece un ejemplo en el que se identifican los cuatro componentes

Figura 2
Ejemplo de instrucción con sus cuatro componentes: instrucción, contexto, datos de entrada e indicador de salida



Fuente: este ejemplo de *prompt* se generó utilizando el modelo GPT-4o de OpenAI. La figura destaca sus partes esenciales de un *prompt* (octubre de 2024, <https://bit.ly/3V1bKdk>).

Dicho esto, también debe reconocerse que el *prompt*, a través de sus elementos, se manifiesta en última instancia como una forma de expresión de datos (por ejemplo, palabras, *tokens*, instrucciones). Entonces, también puede entenderse como una cadena de caracteres que no son diferentes de las frases que escribo en este computador, las fotos que veo o la última película que vi en Netflix la semana pasada.

Hong Liu (2014) sostiene que, para muchos, los datos son simplemente aquello que se percibe, los bloques de construcción que constituyen nuestro conocimiento de la realidad. Desde esta perspectiva, la realidad se concibe como algo externo e independiente de la conciencia. Los datos, a su vez, suelen reducirse a números e información que reflejan hechos objetivos. Sin embargo, como subraya Hong Liu (2014), los datos no son solo un hecho objetivo (impulsos eléctricos convertidos en ceros y unos), sino el resultado de la actividad cognitiva humana, una actividad caracterizada por el reflejo subjetivo de fenómenos objetivos:

Los datos se originan en la observación y las mediciones, demostrando las características intrínsecas de las cosas, cuya objetividad sigue estando bajo la influencia de los factores subjetivos humanos, y esta propiedad va acompañada de datos incluso desde el origen de los números. Los datos son, por así decirlo, un vínculo para conectar la objetividad de las cosas y la subjetividad de las cosas humanas, y también son un puente para que los seres humanos reconozcan el mundo (p. 65).

154



En este examen inicial del *prompt*, este ha sido considerado únicamente como un elemento material, que comparte características comunes con otras formas de datos. A este nivel, sigue siendo un lenguaje sin rostro, un discurso sin voz y una máquina sin vida.

Esta perspectiva es incompleta si no se reconoce que, a diferencia de las imágenes o del texto que escribo, el *prompt* posee una dimensión interactiva y funcional: no es una simple información pasiva, sino un vehículo intencional que desencadena un proceso computacional. En otras palabras, el *prompt* no es solo una representación material de datos, sino una interfaz entre el usuario y el modelo. Al ser construido a propósito para producir un resultado específico del sistema, el *prompt* introduce una dimensión pragmática y técnica que a menudo permanece imperceptible para el usuario. Cuando se examinan con mayor detalle, estos parámetros —las herramientas técnicas que regulan el rendimiento del modelo— revelan una convergencia entre la intención humana, la mediación tecnológica y la construcción de significado.

En la siguiente sección se examinan tres parámetros clave que influyen en el resultado de las interacciones mediante *prompts*, incluidos elementos que a menudo permanecen implícitos y no son reconocidos por los usuarios. Aunque el enfoque se limita a estos casos seleccionados, es fundamental reconocer que un análisis exhaustivo requeriría abordar aspectos adicionales más allá del ecosistema de OpenAI los mensajes del

sistema y la asignación de roles). Se ilustrará cómo la configuración de estos parámetros conlleva implicaciones epistemológicas, éticas y existenciales, invitando a reflexionar sobre el impacto más amplio del diseño de *prompts*. La deconstrucción de estos elementos revela tensiones subyacentes entre precisión y creatividad, reducción y apertura, orden y fluidez, y expone tanto los límites técnicos de los *chatbots* como las profundas apuestas filosóficas de la intervención humana en este ámbito.

Exploración de los parámetros clave del prompt

LA TEMPERATURA: ENTRE PRECISIÓN Y CREATIVIDAD

El parámetro “temperatura” regula la aleatoriedad de las respuestas del modelo, influyendo en cómo selecciona las palabras o *tokens* al generar una salida. Los valores bajos suelen producir resultados más deterministas, mientras que los altos introducen diversidad y creatividad.

Los estudios recientes han analizado cómo afectan los ajustes de temperatura a la creatividad y la resolución de problemas en LLM. Aunque a menudo se le denomina “parámetro de creatividad”, sus efectos parecen ser más complejos (Peeperkorn *et al.*, 2024). Los valores más altos tienden a aumentar la novedad, pero también pueden introducir cierto grado de incoherencia, sobre todo en tareas como la generación de narraciones.

Específicamente en la ingeniería de *software*, las variaciones en los parámetros de entrada, incluida la temperatura, pueden mejorar significativamente el rendimiento de la generación de código, aunque lograr un control óptimo sigue siendo un reto (Döderlein *et al.*, 2023). Estos hallazgos sugieren que el papel de la temperatura en el rendimiento de los LLM es complejo y altamente dependiente del contexto. En concordancia con quienes abordan este parámetro con cautela, se enfatiza que no se trata únicamente de la temperatura, sino también de cómo otros valores influyen en el resultado del *prompt*.

Por ejemplo, cuando se le solicita a ChatGPT que responda a la pregunta “¿Qué es la Luna?” (Tabla 2), ajustando únicamente el parámetro de temperatura y manteniendo el resto de la configuración sin cambios, observaríamos que las respuestas varían no solo en su estructura, sino también en su significado.

Tabla 2
Respuestas del modelo a la pregunta: "¿Qué es la Luna?"

Temperatura	Respuesta
0,2	La Luna es el único satélite natural de la Tierra, orbita alrededor del planeta e influye en las mareas, con una superficie marcada por cráteres y cuencas.
1	La Luna es el satélite natural de la Tierra, un cuerpo celeste que orbita alrededor de nuestro planeta y es su vecino cósmico más cercano.

Fuente: respuestas generadas por GPT-4o (OpenAI) en dos configuraciones distintas de temperatura: 0,2 y 1,0. El experimento se realizó en el entorno OpenAI Playground. Para consultar el código utilizado en este experimento véase el apéndice A.

156



A temperaturas más bajas (0,2), la respuesta es precisa, objetiva y técnica, centrándose en detalles como el papel de la Luna al orbitar la Tierra e influir en las mareas y sus características físicas, como cráteres y cuencas. Esto refleja la tendencia del modelo a dar prioridad a la claridad, la precisión y el determinismo cuando la aleatoriedad es mínima.

En cambio, a mayor temperatura (1,0) la respuesta adopta un enfoque más expresivo y estilizado, describiendo la Luna como “cuerpo celeste” y “vecino cósmico más cercano”. Aunque sigue siendo coherente y pertinente, esta versión introduce un tono más poético y creativo, ofreciendo una perspectiva más amplia y menos técnica. Esta segunda definición presenta la Luna como un objeto físico y utiliza recursos expresivos que evocan imágenes más sugerentes.

A pesar de estar limitado a una sola frase, el ajuste de la temperatura influye en el nivel de detalle, la creatividad y la variabilidad de las respuestas, lo que demuestra la flexibilidad del modelo para adaptarse a distintos estilos y requisitos de salida.

Los parámetros de configuración no son neutros; responden a la intencionalidad humana orientada hacia distintas formas de ser y conocer. Optar por la precisión refleja una búsqueda de objetividad y control, subordinando el lenguaje a criterios pragmáticos y funcionales. Por el contrario, privilegiar la creatividad abre el horizonte a respuestas menos previsibles, permitiendo que surjan nuevas posibilidades de sentido. Se debe evitar la tentación de asignar valor moral a una opción sobre la otra. Sin embargo, lo que sí puede afirmarse es que las variaciones sintácticas y semánticas influyen claramente en la modelización de las distintas narraciones.

De ahí se derivan innumerables problemas de investigación: al configurar el modelo para “ser creativo”, proyecto mi libertad sobre un

sistema que no la posee, reduciendo la creatividad a una operación probabilística: ¿qué implica esta búsqueda de precisión y creatividad para la naturaleza del lenguaje? Al confiar en un sistema que emula, pero no comprende el significado, ¿no corremos el riesgo de empobrecer la auténtica imprevisibilidad de la experiencia humana? Entonces, ¿qué peso tienen estas consideraciones a la hora de elaborar modelos concretos? Por ejemplo, más allá del debate técnico y de los diversos estudios de casos en busca de optimización, ¿Qué debe entenderse por temperatura en contextos donde el objetivo es desarrollar un chatbot educativo o uno diseñado para cumplir una función social específica? Lamentablemente, no existen respuestas definitivas a estas preguntas. Por desgracia, aún no tenemos respuestas definitivas a estas preguntas. Normalmente, estos parámetros se ajustan empíricamente, modulando los valores en función de los resultados prácticos. Por lo tanto, se necesita complementar el enfoque técnico predominante con una intervención más rigurosa y crítica de la filosofía y las ciencias sociales.

TOP P: LA PROBABILIDAD COMO HORIZONTE DE SENTIDO

El segundo ejemplo examina el parámetro Top P. En este caso, el parámetro determina qué probabilidades de *token* se tendrán en cuenta en la salida, orientando el modelo hacia respuestas más predecibles (valores bajos) o salidas más diversas y menos esperadas (valores altos). Este parámetro, también conocido como “muestreo de núcleos”, es una técnica que controla la probabilidad acumulada de las palabras seleccionadas por un modelo lingüístico al generar una respuesta.

Al ajustar el Top P, se decide cuántas palabras o *tokens* (en función de su probabilidad) se tendrán en cuenta en el proceso de generación de texto. El ejemplo (tabla 3) muestra los resultados de la consulta con diferentes valores de Top P.

Tabla 3
Respuestas del modelo al prompt: En un máximo de cuatro frases, describa una escena de una "ciudad futurista" con diferentes configuraciones Top P

Top P	Respuesta
0,2	En el corazón de la ciudad futurista, se extienden altísimos rascacielos con relucientes fachadas de cristal hacia el cielo, y sus superficies vivas tienen dinámicos anuncios holográficos. Los drones autónomos están en el aire, entregando paquetes y vigilando en tiempo real, mientras

	que los vehículos sin conductor se deslizan silenciosamente por las autopistas de varios niveles. Las calles están repletas de personas, muchas de ellas acompañadas por robots controlados por IA que conversan con ellos o los ayudan en sus tareas cotidianas. En lo alto, una red de espacios verdes y paneles solares contrasta con el área urbana y ofrece un atisbo de naturaleza en medio de la maravilla tecnológica.
1	Las luces de neón bañan la ciudad de colores vibrantes mientras los vehículos autónomos se deslizan silenciosamente por las autopistas de varios niveles. Se ven rascacielos imponentes adornados con jardines verticales, y sus superficies acristaladas reflejan la bulliciosa metrópolis. La gente, acompañada de robots humanoides, pasea entre muestras de arte holográfico en grandes plazas con paneles solares. Por encima, los drones vuelan en el aire, gestionando eficazmente las entregas y la vigilancia, al tiempo que garantizan el perfecto funcionamiento de la ciudad.

Fuente: respuestas generadas por GPT-4o (OpenAI) con dos ajustes Top P diferentes: 0,2 y 1,0. El experimento se realizó en el entorno OpenAI Playground. Para consultar el código utilizado en este experimento véase el apéndice A.

158



Como puede verse, en Top P = 0,2, la respuesta es más estructurada y lineal, progresando metódicamente de las calles al cielo, centrándose en detalles tecnológicos y descripciones funcionales. La sintaxis en este ejemplo es específica. Todo el párrafo es una amplia descripción de objetos y cosas. Las frases se basan en expresiones subordinadas, creando un flujo descriptivo constante. El lenguaje hace hincapié en la precisión, destacando ideas como “autopistas de varios niveles” y “espacios verdes”, que reflejan un equilibrio entre tecnología y naturaleza.

En cambio, en Top P = 1,0, la respuesta es más dinámica y segmentada, centrándose en imágenes vívidas y estilísticas más que en una progresión espacial estricta. La sintaxis es más breve, lo que confiere al texto un ritmo más rápido. El lenguaje es más rico y evocador, introduciendo expresiones creativas como “luces de neón”, “jardines verticales” y “exposiciones de arte holográfico”, que transmiten un paisaje urbano visualmente imaginativo.

En resumen, el Top P = 0,2 produce una descripción técnica detallada, mientras que el Top P = 1,0 produce una representación más creativa, expresiva y visualmente rica. Se puede llegar a decir que el futuro se ve de forma diferente en ambos casos. En definitiva, en una descripción, es el futuro desde el punto de vista de las cosas; en el segundo ejemplo, es el futuro desde el punto de vista subjetivo.

El hecho de delimitar las opciones probables es, en definitiva, una selección que afecta al significado. Por otra parte, también equivale a establecer filtros intencionales en el horizonte lingüístico del modelo. Sin embargo, esta decisión no es neutra: es una elección discursiva, incluso

en términos foucaultianos. Según la estrategia, las voces potenciales y las interpretaciones, cada afirmación excluye o incluye. ¿Qué realidades silenciamos al restringir la probabilidad? ¿Qué posibilidades de lenguaje y pensamiento se pierden cuando priorizamos la precisión técnica sobre la apertura ontológica del discurso?

Michel Foucault (1994a, p. 241) creía que las prácticas discursivas no son pura y simplemente formas de crear discurso. Toman forma en el conjunto de técnicas, instituciones y pautas de comportamiento, entre otras, que las imponen y sostienen. De ahí que es de gran importancia centrar el análisis en estos filtros inmediatos. Por un lado, pone de relieve su conexión con narrativas más generales; por otro, insiste en la falta de neutralidad de la tecnología. Al igual que desde la fenomenología crítica se asume que no hay fenómeno fuera de las circunstancias políticas tampoco Foucault (1994b, p. 465) pensó que el poder estuviera fuera del discurso. El poder no es ni la fuente ni el origen del discurso, sino que funciona a través de él.

***PENALIZACIÓN DE FRECUENCIA Y PENALIZACIÓN DE PRESENCIA:
LA REPETICIÓN COMO FENÓMENO DE SENTIDO***

El parámetro “penalización de frecuencia” está diseñado para reducir la repetición de palabras o frases en el texto generado, y funciona añadiendo un valor a la probabilidad logarítmica de un *token* cada vez que aparece en el resultado generado. En consecuencia, cuanto mayor sea el valor de la penalización por frecuencia, más evitará el modelo reutilizar los mismos *tokens*, aumentando la variedad y reduciendo la redundancia. Por ejemplo, con un valor bajo, el modelo podría repetir palabras relevantes para el contexto, mientras que, con un valor alto, buscará alternativas para mantener la diversidad lingüística.

Por otro lado, el parámetro de “penalización por presencia” promueve la inclusión de una mayor variedad de *tokens*, fomentando la creatividad en el texto generado. Funciona restando un valor a la probabilidad logarítmica de un *token* cada vez que se utiliza, lo que disminuye la probabilidad de que el modelo repita palabras generadas previamente. Un valor alto de penalización por presencia aumenta la posibilidad de introducir nuevas ideas o conceptos, evitando que el modelo se limite a los mismos temas.

Tabla 4
Respuestas del modelo a la pregunta:
"Escribe una breve descripción de un bosque en primavera
en no más de tres frases" con diferentes configuraciones
de penalización por frecuencia y presencia

Frecuencia y Penalización	Sanción por presencia	Respuesta
0	0	En primavera, el bosque se llena de vida con un tapiz de verdes vibrantes y hojas frescas que se despliegan por todas las ramas. El aire se llena con la sinfonía del canto de los pájaros y el suave susurro del nuevo follaje. Las flores silvestres salpican el suelo del bosque, proporcionando coloridas manchas en medio del exuberante verdor.
1	0	En primavera, el bosque cobra vida con un vibrante tapiz de hojas verdes frescas y flores, mientras el sol se filtra suavemente a través del dosel. El aire sopla con energía renovada, lleno del canto de los pájaros que regresan de sus migraciones invernales y del suave susurro de la fauna que se despierta. Las flores silvestres cubren el suelo del bosque de brillantes tonalidades, transformándolo en un colorido mosaico que celebra los nuevos comienzos.
0	1	En primavera, el bosque despierta con un vibrante tapiz de verdes a medida que las hojas se despliegan de los árboles. Una sinfonía de cantos de pájaros llena el aire, mientras las flores silvestres cubren el suelo del bosque y lo llenan de color. El fresco aroma de la tierra se mezcla con la luz del sol, creando una atmósfera de renovación y vitalidad.
1	1	En primavera, el bosque se llena de vida con un vibrante dosel de hojas verdes y coloridas flores silvestres que cubren el suelo. El aire se llena con el canto de los pájaros y el suave susurro del nuevo follaje agitado por una suave brisa. La luz del sol se cuele entre los árboles y dibuja alegres dibujos en el suelo del bosque mientras la naturaleza despierta de su letargo invernal.

Fuente: respuestas generadas por GPT-4o (OpenAI) bajo diferentes configuraciones de penalización por frecuencia y penalización por presencia. El experimento se llevó a cabo en el entorno OpenAI Playground. Para consultar el código utilizado en este experimento véase el apéndice A.

Las respuestas muestran cómo la penalización por frecuencia y presencia influye en el estilo, la estructura y la variedad del texto generado. En el primer caso, en el que ambas penalizaciones se fijan en cero, la respuesta es coherente pero repetitiva y se basa en gran medida en la elección de palabras predecibles. Con una penalización de frecuencia más alta, como en el segundo caso, el modelo reduce la redundancia, produciendo un resultado más diverso y refinado a la vez que mantiene la



atención en la descripción central. Cuando se aumenta la penalización por presencia, como en el tercer caso, la respuesta se vuelve más creativa, introduciendo nuevas imágenes y detalles sensoriales que amplían el abanico temático. Por último, cuando se aplican ambas penalizaciones simultáneamente, la respuesta equilibra la variedad y novedad, evitando la repetición y potenciando la expresividad.

Por tanto, el ajuste de estos parámetros implica un compromiso entre la uniformidad algorítmica y la riqueza del lenguaje. En última instancia, ¿cuál es el límite entre riqueza y eficacia? ¿Cuánto debe sacrificar un usuario en aras de la optimización? ¿En qué contextos es ideal la precisión y en cuál es la profundidad semántica? ¿Hasta qué punto debería el lenguaje reflejar la diferencia en lugar de la repetición mecánica del mundo?

Estas son solo algunas de las muchas preguntas que pueden plantearse y dejarse para futuras investigaciones. La ingeniería de *prompt* puede basarse en otros parámetros que no he mencionado aquí producto de la brevedad. Entre ellos están la “longitud máxima”, que limita el número de *tokens* generados, y las “secuencias de parada”, que indican al modelo cuándo debe detener su salida. También hay que prestar atención al “mensaje del sistema”, un elemento fundamental que influye directamente en la interacción y el comportamiento del modelo.

Conclusiones

A partir del análisis de distintos parámetros, la filosofía del *prompting* se perfila como una herramienta indispensable para abordar cuestiones relativas a las estructuras ontológicas, éticas y epistemológicas. Los resultados indican que las indicaciones no son simples instrucciones técnicas, sino mediaciones impregnadas de intencionalidad humana.

Como se ha comentado anteriormente, el *prompt* funciona como un espacio de encuentro entre la intencionalidad del sujeto y la respuesta generada por el modelo. Su análisis fenomenológico permite comprender cómo la interacción con los *chatbots* introduce nuevos modos de ser en el mundo. Aunque el lenguaje técnico suele ocultar su complejidad, el presente estudio ha demostrado que cada *prompt* refleja decisiones humanas, explícitas o implícitas, que configuran los límites y las posibilidades del conocimiento.

Lo que es crucial en este caso es que, hoy en día, la realidad del *prompt* es casi omnipresente. Son pocos los espacios en los que los individuos no se enfrentan a *chatbots* que han sido configurados mediante *prompts* para ofrecer respuestas específicas y reforzar determinadas narra-

tivas del mundo, o bien que son utilizados mediante prácticas de *prompting* rudimentarias. El resultado apunta en la misma dirección en todos los casos: se necesita más conocimiento y educación sobre estas herramientas.

El análisis de las instrucciones revela que la responsabilidad sigue cayendo en el agente humano. Es el usuario quien define los parámetros, estructura las instrucciones y asigna significado a las respuestas. Por lo que es imperativo reconocer esta responsabilidad para evitar caer en narrativas que antropomorficen a la IA o exageren su alcance. En este sentido, los parámetros de configuración como la temperatura, el Top P y las penalizaciones por frecuencia y presencia demuestran cómo las elecciones técnicas conllevan a implicaciones tanto discursivas como éticas. La manipulación de estos valores no es neutral: privilegiar la precisión sobre la creatividad, o viceversa, determina los resultados del modelo y la construcción de significados.

Por último, este trabajo propone la necesidad de un enfoque interdisciplinario que integre la filosofía crítica con el desarrollo técnico de la IA. La ingeniería de prompts debería ser capaz de incorporar un análisis experimental del *prompting* junto con un análisis crítico del mismo. A su vez, la educación tiene un papel clave en la formación de individuos capaces de relacionarse críticamente con los sistemas de IA, comprendiendo tanto sus alcances como sus límites.

No se trata de que cada línea conlleve a una reflexión filosófica, sino de que la filosofía debería pensar mucho más en las condiciones de posibilidad del código, en las estrategias y técnicas discursivas de un lenguaje que empieza a ser cada vez más común, porque en definitiva, como explica Vallor (2024), esta revolución no es más que una ilusión, un gigantesco espejo “hecho de código, construido para consumir nuestras palabras, nuestras decisiones, nuestro arte, nuestras expresiones de amor y cuidado, solo para reflejarlas de nuevo hacia nosotros” (p. 194). Esto significa que, en última instancia, los problemas de esta nueva era radican mucho más en nosotros de lo que pensamos.

Bibliografía

- BORY, Paolo, NATALE, Simone, & KATZENBACH, Christian
2024 Strong and Weak AI Narratives: An Analytical Framework. *AI & Society*, 40, 2107-2117. <https://doi.org/10.1007/s00146-024-02087-8>
- BRINGSJORD, Selmer, & GOVINDARAJULU, Naveen Sundar
2024 Artificial Intelligence. En E. N. Zalta & U. Nodelman (eds.), *The Stanford Encyclopedia of Philosophy*. Metaphysics Research Lab, Stanford University. <https://bit.ly/4BcdZp>



- BUCKNALL, Benjamin, & DORI-HACOHEN, Shiri
 2022 Current and Near-Term AI as a Potential Existential Risk Factor. *Proceedings of the 2022 AAAI/ACM Conference on AI, Ethics, and Society*, 119-129. <https://doi.org/10.1145/3514094.3534146>
- CHIARELLA, Salvatore, TORROMINO, Giulia, GAGLIARDI, Dionigi, ROSSI, Dario, BABILONI, Fabio, & CARTOCCI, Giulia
 2022 Investigating the Negative Bias towards Artificial Intelligence: Effects of prior Assignment of AI-authorship on the Aesthetic Appreciation of Abstract Paintings. *Computers in Human Behavior*, 137, 107406. <https://doi.org/10.1016/j.chb.2022.107406>
- DESPANDE, Ameet, RAJPUROHIT, Tanmay, NARASIMHAN, Karthik, & KALYAN, Ashwin
 2023 Anthropomorphization of AI: Opportunities and Risks. *Arxiv*. 23 de mayo. <https://doi.org/10.48550/ARXIV.2305.14784>
- DÖDERLEIN, Jean-Baptiste, ACHER, Mathieu, KHELLADI, Djamel, & COMBE-MALE, Benoit
 2023 Piloting Copilot and Codex: Hot Temperature, Cold Prompts, or Black Magic? *Arxiv*. 15 de febrero. <https://doi.org/10.48550/arXiv.2210.14699>
- FLORIDI, Luciano
 2023 *The Ethics of Artificial Intelligence*. Oxford University Press. <https://doi.org/10.1093/oso/9780198883098.001.0001>
- FOUCAULT, Michel
 1994a *Dits et Écrits 1970-1975* (vol. 2). París: Gallimard.
 1994b *Dits et Écrits 1976-1979* (vol. 3). París: Gallimard.
- GOERTZEL, Ben
 2015 Superintelligence: Fears, Promises and Potentials: Reflections on Bostrom's Superintelligence, Yudkowsky's From AI to Zombies, and Weaver and Veitas's "Open-Ended Intelligence". *Journal of Ethics and Emerging Technologies*, 25(2), 2. <https://doi.org/10.55613/jeet.v25i2.48>
- GONZÁLEZ, Jorge
 2024 Artificial Intelligence and the New Dynamics of Social Death: A Critical Phenomenological Inquiry. SSRN. <http://dx.doi.org/10.2139/ssrn.4955995>
- GU, Jindong, HAN, Zhen, CHEN, Shuo, BEIRAMI, Ahmad, HE, Bailan, ZHANG, Gengyuan, LIAO, Ruotong, QIN, Yao, TRESP, Volker, & TORR, Philip
 2023 A Systematic Survey of Prompt Engineering on Vision-Language Foundation Models. *Arxiv*. 24 de julio. <https://doi.org/10.48550/arXiv.2307.12980>
- HARARI, Yuval Noah
 2023 Yuval Noah Harari argues that AI has hacked the operating system of human civilisation. *The Economist*. 28 de abril. <https://bit.ly/3DippkO>
- HARTMANN, Jochen, SCHWENZOW, Jasper, & WITTE, Maximilian
 2023 *The political ideology of conversational AI: Converging evidence on ChatGPT's pro-environmental, left-libertarian orientation*. *Arxiv*. 5 de enero. <https://doi.org/10.48550/arXiv.2301.01768>
- HEIDEGGER, Martin
 1996 *Being and Time*. Nueva York: State University of New York Press.
- KUMAR, Shailendra, & CHOUDHURY, Sanghamitra

- 2023 Humans, Super Humans, and Super Humanoids: Debating Stephen Hawking's doomsday AI forecast. *AI and Ethics*, 3(3), 975-984. <https://doi.org/10.1007/s43681-022-00213-0>
- LIU, Hong
2014 Philosophical Reflections on Data. *Procedia Computer Science*, 30, 60-65. <https://doi.org/10.1016/j.procs.2014.05.381>
- MERLEAU-PONTY, Maurice
1993 *Fenomenología de la percepción*. Buenos Aires: Planeta.
- MESKÓ, Bertalan
2023 Prompt Engineering as an Important Emerging Skill for Medical Professionals: Tutorial. *Journal of Medical Internet Research*, 25, e50638. <https://doi.org/10.2196/50638>
- PEEPERKORN, Max, KOUWENHOVEN, Tom, BROWN, Dan, & JORDANOUS, Anna
2024 Is Temperature the Creativity Parameter of Large Language Models? *Arxiv*. 1 de mayo. <https://doi.org/10.48550/arXiv.2405.00492>
- PLACANI, Adriana
2024 Anthropomorphism in AI: Hype and fallacy. *AI and Ethics*, 4(3), 691-698. <https://doi.org/10.1007/s43681-024-00419-4>
- PROMPTING TECHNIQUES-NEXTRA
2024 *Prompting Techniques*. <https://bit.ly/4gfQaFc>
- RICOEUR, Paul
2010 *Memory, History, Forgetting*. University of Chicago Press.
- ROSER, Max
2024 The Brief History of Artificial Intelligence: The World has Changed Fast-What Might be Next? *Our World in Data*. 6 de diciembre. <https://bit.ly/4iE1sVq>
- RUSSELL, Stuart, & NORVIG, Peter
2020 *Artificial Intelligence-A Modern Approach*. Englewood Cliffs, NJ: Pearson.
- SARTRE, Jean-Paul
1984 *Being and Nothingness*. Washington Square Press.
- SEARLE, John
1980 Minds, Brains, and Programs. *The Behavioral and Brain Sciences*, (3), 417-457. <https://doi.org/10.1017/S0140525X00005756>
- SILVA, Altieres de Oliveira, & JANES, Diego dos Santos
2022 The Emergence of ChatGPT and its Implications for Education and Academic Research in the 21st Century. *Review of Artificial Intelligence in Education*, 3, e6-e6. <https://doi.org/10.37497/rev.artif.intell.educ.v3i00.6>
- VALLOR, Shannon
2024 *The AI Mirror: How to Reclaim Our Humanity in an Age of Machine Thinking* (1ª ed.). Oxford University Press New York. <https://doi.org/10.1093/oso/9780197759066.001.0001>
- WEISS, Gail, MURPHY, Ann, & SALAMON, Gayle
2019 *50 Concepts for a Critical Phenomenology*. Northwestern University Press. <https://doi.org/10.2307/j.ctvmx3j22>
- ZARIFHONARVAR, Ali
2023 *Economics of ChatGPT: A Labor Market View on the Occupational Impact of Artificial Intelligence* [Ponencia]. Kiel, Hamburg: ZBW-Leibniz Information Centre for Economics. <https://bit.ly/4gBn4Qa>



Apéndice A: datos complementarios

Los *scripts* de Python y los datos complementarios utilizados para realizar los experimentos presentados en este artículo se encuentran en el Open Science Framework (OSF) y se puede acceder a ellos en <https://doi.org/10.17605/OSF.IO/UA3E5>.

Este archivo incluye:

1. *Scripts* de Python para experimentos de temperatura, muestreo Top P y parámetros de penalización.
2. Un archivo README que detalla el propósito de cada *script*, cómo utilizarlos y sus dependencias.
3. Un archivo LICENSE que especifica que la obra se publica bajo la Licencia Universal Creative Commons Zero v1.0 (CC0), que permite su uso, modificación y distribución sin restricciones.

Además, se ha cargado junto al artículo un archivo ZIP que contiene los *scripts* de Python para facilitar el acceso.



Declaración de Autoría - Taxonomía CRediT	
Autor/es	Contribuciones
Jorge González Arocha	Conceptualización; metodología; investigación; redacción-borrador original; redacción-revisión y edición; visualización; software; validación.

Declaración de uso de inteligencia artificial
El autor Jorge González Arocha DECLARA que el manuscrito “Fenomenología crítica del prompting en la inteligencia artificial” SÍ INCLUYE el uso de inteligencia artificial (IA). La IA se empleó exclusivamente para generar ejemplos de prompts y sus respuestas mediante la API de OpenAI, lo cual se requería para desarrollar el análisis propuesto en el artículo. Algunas de las tablas y figuras resultantes forman parte esencial de la investigación y no habrían podido realizarse sin esta herramienta.

Fecha de recepción: 15 de diciembre de 2024

Fecha de revisión: 25 de febrero de 2025

Fecha de aprobación: 20 de abril de 2025

Fecha de publicación: 15 de julio de 2025