Prompts efectivos para ChatGPT

CÓRDOBA CREA FUTURO

En el panorama laboral actual, la inteligencia artificial (IA) generativa, y específicamente los modelos de lenguaje grandes (large language models, LLM) como ChatGPT, han dejado de ser una novedad tecnológica para convertirse en una herramienta fundamental de productividad. Sin embargo, como ocurre con cualquier herramienta poderosa, su valor no reside en su mera existencia, sino en la habilidad para operarla con destreza.

Aquí surge una distinción clave para la empleabilidad: la diferencia entre un usuario de IA y un operador de IA. Un usuario es pasivo; trata al modelo como un buscador mejorado, haciendo preguntas simples y esperando respuestas. Un operador, en cambio, es un profesional activo que diseña, prueba y refina instrucciones de manera deliberada para obtener resultados predecibles, eficientes y de alta calidad.

El mercado laboral no recompensará, a largo plazo, a quienes simplemente «usen» IA, sino a los operadores: aquellas personas que pueden dirigir la herramienta para crear valor tangible. La llamada ingeniería de instrucciones o prompt engineering es la disciplina que formaliza esta habilidad. Se trata de una metodología estructurada para optimizar la interacción con la IA y lograr resultados precisos (Velásquez-Henao et al., 2023).

- Fundamentos de un prompt
- Buenas prácticas para la generación de prompts
- Alucinaciones: ¿qué son y cómo manejarlas?
- Conclusión
- Referencias
- Descarga en PDF

Fundamentos de un prompt

En su forma más simple, un prompt es un texto en lenguaje natural (una pregunta, una orden, una declaración) que se proporciona a un modelo de inteligencia artificial para guiarlo en la generación de una respuesta (Iyer, 2025). Es la principal interfaz que tenemos para comunicarnos con el modelo y recuperar la información que buscamos.

Sin embargo, un prompt efectivo es mucho más que una simple pregunta; es un «paquete de intención» cuidadosamente construido (Kinnoch, s. f.). La calidad de la respuesta del modelo es directamente proporcional a la claridad con la que el operador estructura y transmite esa intención.

Como se explicó en módulos anteriores, los modelos de lenguaje no «entienden» el significado como lo haría una persona; generan texto en función de patrones aprendidos a partir de grandes volúmenes de datos. Una pequeña

variación en cómo formulamos la solicitud puede marcar la diferencia entre una respuesta irrelevante y una que supere nuestras expectativas (Federiakin et al., 2024). Precisar claramente lo que queremos ayuda al modelo a orientarse y generar resultados más acertados.

Si bien un prompt puede ser tan simple como «¿cuál es la capital de Francia?», los prompts de nivel profesional, utilizados en entornos laborales, suelen tener una estructura más compleja que incluye, por lo general, cuatro componentes:

La instrucción

Es la tarea específica y directa que se le solicita al modelo. Por ejemplo, «resume...», «traduce...», «analiza...», «escribe código...».

El contexto

Corresponde a la información de fondo que enmarca la tarea. Ayuda al modelo a comprender el propósito de la solicitud, el público objetivo o las circunstancias que la rodean. Por ejemplo, «eres un analista financiero preparando un informe para inversores...».

Los datos de entrada —

Se refiere al texto, los datos o la información concreta sobre los que el modelo debe trabajar. Por ejemplo, «basado en este correo electrónico del cliente: [texto]...».

El indicador de salida —

Establece especificaciones claras sobre el formato, la longitud, el tono o la estructura deseada de la respuesta. Por ejemplo, «...en tres viñetas», «...en formato de tabla Markdown», «...en menos de 50 palabras».

Prompt engineering

El *prompt engineering* (a veces traducido como ingeniería de instrucciones o diseño de *prompts*) es la disciplina y metodología que permite diseñar, estructurar, probar y refinar *prompts* para modelos de lenguaje. Su objetivo es guiar de forma efectiva a un modelo de lenguaje grande (LLM) para que genere la salida deseada, asegurando que sea precisa, pertinente y adecuada al contexto.

Esta disciplina resulta necesaria porque, aunque los LLM son herramientas versátiles, su uso sin orientación puede derivar en respuestas inexactas o contextualmente inapropiadas (Velásquez-Henao et al., 2023).

El *prompting* no es un proceso de un solo intento (*one-shot*), sino un proceso iterativo. Una metodología formal para este proceso incluye los siguientes elementos:

- **Definir el objetivo.** ¿Qué necesito exactamente?
- **Diseñar el prompt:** construir la instrucción inicial.
- **Evaluar la respuesta:** ¿es esto lo que pedí? ¿Es precisa?
- **Iterar:** refinar el prompt con base en la evaluación para mejorar el resultado.

CONTINUAR

Buenas prácticas para la generación de prompts

A continuación, se presentan una serie de técnicas y recomendaciones probadas para mejorar la calidad de tus prompts, cada una acompañada de ejemplos prácticos.

Establecer el contexto

El contexto es la técnica más importante para reducir la ambigüedad. El modelo no sabe nada sobre ti, tu trabajo, tus objetivos ni tu audiencia. Proporcionar esta información de fondo —es decir, el «por qué» de la tarea— es clave para obtener una respuesta relevante.

Por ejemplo, si estamos organizando una reunión de ventas, el resultado puede variar mucho dependiendo de cuánta información le proporcionemos al modelo:

- **Prompt débil.** «Dame ideas para la reunión de ventas.»

- **Prompt fuerte (con contexto):** «Estoy preparando la reunión trimestral de ventas (Q3) para mi equipo. Mi rol es gerente de ventas y mi audiencia está compuesta por diez personas de mi propio equipo. El objetivo de la reunión es motivarlos, no criticarlos. Necesito tres ideas clave para una presentación de PowerPoint que celebre los logros del segundo trimestre y establezca metas realistas para el cuarto trimestre».

Ser específico y detallado

Los prompts deben ser lo más específicos y detallados posible respecto a lo que se desea obtener. Incluir información concreta ayuda a reducir la ambigüedad y orienta al modelo con mayor precisión hacia lo que necesitas. Si tu solicitud es vaga, la respuesta tenderá a ser genérica. En cambio, al delimitar claramente el tema, la extensión, el formato o los puntos clave a cubrir, aumentan las probabilidades de recibir una respuesta útil y precisa (MIT Sloan Teaching & Learning Technologies, s. f.).

Imaginemos que solicitamos consejos de marketing para un negocio pequeño. Las diferencias entre un prompt genérico y uno específico son evidentes:

- **Prompt genérico.** «Dame consejos de marketing digital.»
- **Prompt específico:** «brinda cinco consejos prácticos de marketing digital enfocados en redes sociales para una pequeña empresa de comida orgánica. Por favor, incluye estrategias de bajo costo y menciona ejemplos de campañas exitosas recientes. Responde en un tono motivador y en formato de lista».

En el segundo caso, se han añadido detalles clave: cantidad de elementos, ámbito (redes sociales), tipo de negocio (comida orgánica), restricciones (bajo costo), y formato y tono deseado. Con esas instrucciones claras, el modelo podrá generar una respuesta mucho más útil, estructurada y pertinente.

De hecho, la precisión en el prompt suele traducirse en mayor relevancia en la respuesta: la granularidad —es decir, el grado de detalle de tu pregunta— está directamente relacionada con la utilidad del resultado (MIT Sloan Teaching & Learning Technologies, s. f.). En resumen, al formular un prompt detallado estamos acotando el problema que la IA debe resolver, lo cual se traduce en respuestas más adecuadas y con menos margen de error.

Especificar tono y estilo de respuesta

Los modelos de lenguaje grandes (LLM) son altamente dirigibles, lo que significa que responden muy bien a instrucciones sobre extensión, tono y estilo (Kotha et al., 2025). En un entorno laboral, la forma en que se dice algo es tan importante como el contenido. Por eso, conviene indicarle explícitamente al modelo el tono deseado en cada contexto.

Un estudio centrado en optimizar LLM para la comunicación entre médicos y pacientes con tuberculosis en Argentina instruyó al modelo con pautas culturales y comunicativas específicas para lograr un estilo empático y accesible

(Filienko et al., 2025). Entre las instrucciones utilizadas en ese caso, se incluyen los siguientes ejemplos:

- «Utiliza el "vos" característico del español argentino, pero sin exagerar».
- «Evita tecnicismos médicos; mantén las explicaciones simples y accesibles».
- «Muéstrate siempre empático y dispuesto a escuchar cualquier inquietud».

Este tipo de personalización también es útil en situaciones laborales cotidianas. Por ejemplo, si necesitamos redactar un correo interno, el tono puede marcar una gran diferencia en cómo se recibe el mensaje:

- **Prompt débil.** «Escribe un email sobre la nueva política de trabajo desde casa.»
- **Prompt fuerte (con tono)**: «redacta un borrador de correo para todo el personal (aproximadamente 500 empleados) sobre la nueva política de trabajo híbrido (tres días en oficina, dos en casa). El tono debe ser profesional, pero amigable y empático.

> Queremos que se sienta como una buena noticia que equilibra la flexibilidad y la colaboración, no como una imposición. Evita el lenguaje corporativo rígido como “sinergia” o “imperativo”. El firmante es el director de Recursos Humanos».

Asimismo, puedes solicitar un formato específico: lista de viñetas, párrafo explicativo, diálogo, poema, entre otros. Incluir estas indicaciones dentro del prompt permite alinear la respuesta del modelo con tus necesidades comunicativas (MIT Sloan Teaching & Learning Technologies, s. f.).

Iteración y refinamiento

No siempre obtendrás la respuesta perfecta en el primer intento, y eso es totalmente normal. Una práctica esencial en el uso de modelos de lenguaje es la iteración y el refinamiento: ajustar el prompt en función de las respuestas previas para acercarte progresivamente a lo que necesitas (García-García, 2025).

Conviene pensar la interacción con ChatGPT como un diálogo continuo. Puedes retroalimentar al modelo, corregirlo o afinar tu solicitud tras ver una primera respuesta. Si la respuesta inicial no cumple con tus expectativas, no

dudes en reformular la pregunta. Por ejemplo, quizá en un primer intento escribes «dame un resumen de este texto», y obtienes un resultado demasiado superficial. Podrías refinar así: «el resumen anterior es muy general. Resume nuevamente destacando los tres puntos más importantes y en un máximo de cinco oraciones». Al ofrecer esa retroalimentación, le estás dando al modelo una guía más precisa para mejorar su salida.

Cada interacción representa una oportunidad de aprendizaje, tanto para ti como para ajustar la puntería del modelo. De hecho, los usuarios más experimentados suelen construir el prompt ideal de manera iterativa: comienzan con una versión básica, evalúan la respuesta y luego la ajustan, agregando condiciones, restricciones o detalles adicionales en el siguiente intento. Este proceso de refinamiento continuo permite alcanzar resultados óptimos en pocas rondas de interacción.

En resumen, debemos ser flexibles y pacientes: si la IA no acierta en el primer intento, debemos reformular la solicitud con mayor claridad o detalle. Con el tiempo, desarrollaremos

nuestra propia habilidad en prompt engineering a través de la práctica.

A continuación, se muestra un ejemplo de este proceso de refinamiento aplicado a un caso laboral.

- **Primer prompt:** «dame una descripción de producto para nuestro nuevo software de gestión de proyectos».
- **Primera respuesta de ChatGPT:** genera un texto genérico sobre «aumentar la productividad» y «colaboración fluida».
- **Prompt refinado (con retroalimentación):** «no es lo que busco. Es demasiado genérico y largo. Refínalo para que sea más corto (máximo 50 palabras) y enfócate en nuestro diferenciador clave: la integración con IA para la asignación automática de tareas. El tono debe ser más directo y enérgico, dirigido a gerentes de proyecto ocupados».

Pedir tareas complejas de manera estructurada

Los modelos de lenguaje actuales —especialmente los de última generación— pueden realizar tareas bastante

complejas, pero para ello es fundamental saber cómo formular la solicitud. Cuando necesitas que la IA aborde un problema amplio o con múltiples pasos, una estrategia eficaz consiste en estructurar tu prompt en partes o secciones. Esto ayuda a guiar al modelo a través de la complejidad en lugar de dejarle todo el problema abierto (García-García, 2025).

Por ejemplo, si simplemente dices «escribe un plan de negocio completo», es probable que la respuesta sea demasiado genérica o carezca de foco. En cambio, podrías dividir la tarea en subtareas dentro del mismo prompt:

«Quiero que me ayudes a elaborar un plan de negocio para una cafetería. Primero, propón tres posibles conceptos de cafetería innovadores. Luego, para el concepto elegido, describe el mercado objetivo, la propuesta de valor y una estrategia de marketing resumida. Finalmente, presenta un breve análisis de viabilidad financiera. Responde en secciones separadas con títulos para cada apartado.»

En este caso, la tarea compleja («elaborar un plan de negocio») se ha desglosado en pasos manejables: ideación, análisis de mercado, marketing y finanzas. Esta estructura clara facilita que el modelo entienda que hay múltiples entregables y los aborde uno a uno. Además, al indicar el

formato deseado (secciones con títulos), se incrementa la probabilidad de recibir una respuesta bien organizada y completa.

En general, para solicitar tareas complejas conviene:

- describir con claridad el objetivo general,
- especificar los subobjetivos o componentes que esperas en la respuesta, y
- establecer cualquier formato o criterio de organización.

Los modelos de lenguaje son capaces de procesar instrucciones extensas y multifacéticas, pero lo harán mejor si esas instrucciones están presentadas de forma ordenada y lógica.

Práctica continua y aprendizaje activo

Como toda habilidad, escribir buenos prompts requiere práctica. No basta con leer teoría; la mejora proviene del uso constante de estas herramientas y del descubrimiento gradual de qué tipo de indicaciones funcionan mejor en cada contexto. Te animamos a que, en tus actividades diarias de estudio o trabajo, integres a ChatGPT (u otro modelo) en tus procesos e interés con él.

Por ejemplo, si eres estudiante, puedes formularle preguntas sobre tus apuntes para repasar. Si trabajas en marketing, podrías experimentar con solicitudes de borradores de eslóganes. Si programas, puedes consultarle cómo depurar un error específico. Cada interacción te enseñará algo nuevo sobre cómo la IA interpreta tus instrucciones.

La práctica continua, además, contribuye al desarrollo de la literacidad de prompts: un conjunto de habilidades y conocimientos que te permiten comunicarte de forma eficaz con modelos de lenguaje (Federiakin et al., 2024). Te familiarizarás con los formatos más efectivos, aprenderás qué convenciones suelen funcionar mejor y notarás patrones. Por ejemplo, podrías descubrir que el modelo responde mejor cuando usas viñetas para listar requisitos, o que interpreta de forma ambigua ciertos términos técnicos. Esa sensibilidad mejora con la experiencia.

Investigaciones recientes proponen incluso que la habilidad de prompt engineering debería incluirse en la educación formal, dada su relevancia creciente en múltiples disciplinas (Federiakin et al., 2024).

Recuerda que no existe un prompt perfecto y universal: cada situación es única. Sin embargo, tu experiencia acumulada te permitirá abordar nuevos desafíos con mayor confianza y eficacia.

Role playing: adopción de un rol o persona

Una técnica poderosa para orientar el estilo y la perspectiva de una respuesta es el role playing o asignación de rol. Consiste en indicarle explícitamente al modelo que adopte una identidad, profesión o punto de vista específico al responder (Schulhoff, 2025). De este modo, moldeamos tanto el tono como el contenido de la respuesta para que coincidan con ese rol definido. Por ejemplo, podríamos decir «eres un historiador especializado en la Edad Media...» o «actúa como un entrenador personal experimentado...». Esto

ofrece al modelo un marco de referencia más claro desde el cual generar su salida.

¿Por qué funciona? Los modelos de lenguaje han sido entrenados con grandes volúmenes de texto provenientes de múltiples fuentes (artículos académicos, blogs, foros, libros, etc.). Cuando se les pide que respondan «como si fueran X», internamente buscan patrones lingüísticos y de conocimiento asociados a ese perfil. Esto permite obtener respuestas más coherentes, relevantes y adaptadas al contexto.

De hecho, se ha observado que esta técnica puede mejorar incluso la precisión en algunos casos, ya que al «pensar» como un especialista, el modelo tiende a evitar ambigüedades (Schulhoff, 2025).

Asignar un rol no solo modifica el estilo de la respuesta (formal, técnico, coloquial, etc.), sino que también puede activar conocimientos más específicos. Por ejemplo, si se le indica: «Eres un médico» y luego se formula una consulta sobre síntomas, el modelo responderá utilizando terminología y enfoques propios del ámbito médico.

Proporcionar ejemplos

En lugar de limitarse a describir el formato de salida deseado, es mucho más efectivo mostrarle al modelo exactamente qué tipo de respuesta se espera. Esta técnica se conoce en el ámbito académico como few-shot prompting (aprendizaje con pocos ejemplos) (Kuka, 2025).

Consiste en proporcionar al modelo varios pares de entrada y salida —los llamados shots o ejemplos— para que pueda reconocer el patrón antes de procesar una nueva solicitud. IBM define esta estrategia como el acto de brindarle al modelo «algunos ejemplos de una tarea para guiar su rendimiento», lo cual resulta particularmente útil cuando no se dispone de grandes volúmenes de datos de entrenamiento (Gadesha, s. f.).

Por ejemplo, si necesitas que la IA clasifique el sentimiento en publicaciones breves, podrías usar un prompt como el siguiente:

Prompt (few-shot). Tu tarea es clasificar el sentimiento de los tweets como «Positivo», «Negativo» o «Neutral». Aquí tienes ejemplos del formato deseado:

Tweet: «¡El nuevo software es increíblemente rápido!».

Sentimiento: Positivo

Tweet: «No encuentro el botón de ayuda y estoy atascado».

Sentimiento: Negativo

Tweet: «El webinar está programado para el martes a las 4 p. m.».

Sentimiento: Neutral

Ahora, clasifica este:

Tweet: «He esperado tres días por una respuesta del soporte y sigo sin poder trabajar».

Este enfoque ayuda al modelo a entender no solo la tarea, sino también el formato esperado, lo que mejora la coherencia y precisión de la respuesta.

Interacción invertida

Esta técnica invierte el flujo habitual de la conversación. En lugar de que el operador intente anticipar todo el contexto

que el modelo podría necesitar, se le indica al modelo que formule las preguntas necesarias para recopilar la información que aún falta.

Este enfoque es especialmente útil para tareas ambiguas o cuando el propio operador no está seguro de qué datos son relevantes para la tarea. Veamos un ejemplo.

Prompt (interacción invertida): *«quiero que me ayudes a redactar un plan de negocios para una nueva startup de biotecnología. Es una tarea compleja y no estoy seguro por dónde empezar. Antes de comenzar a redactar, haz todas las preguntas que necesites para entender mi visión, el público objetivo, el presupuesto inicial y nuestros diferenciadores clave. Formularás las preguntas una por una hasta que tengas toda la información que consideres necesaria».*

Con este tipo de prompt, el modelo adopta un rol más activo en la conversación, guiándote para obtener primero los detalles esenciales y, a partir de ellos, generar una respuesta más precisa y útil.

Ajustar el nivel de explicación (ELI – explain like I'm...)

Una habilidad fundamental en cualquier entorno laboral es saber comunicar ideas complejas a audiencias no especializadas. La técnica conocida como ELI (explain like I'm..., o «explícamelo como si yo fuera...»), permite pedirle al modelo que traduzca conceptos difíciles a un nivel de comprensión específico, utilizando analogías o referencias conocidas.

El ejemplo más conocido es ELI5 (explain like I'm 5, o explícamelo como si tuviera 5 años), donde se solicita al modelo que explique un concepto, como «¿qué fue la Revolución Francesa?», en términos simples y accesibles para un niño pequeño, usando imágenes como juguetes, golosinas o juegos.

Esta técnica no solo facilita la comprensión, sino que también ayuda a adaptar el mensaje a distintas audiencias dentro del entorno profesional. Veamos un ejemplo:

Prompt (ELI adaptado): *«tengo que explicar nuestro nuevo "modelo de atribución algorítmica" al equipo de ventas. Explica qué significa como si le hablaras a una persona inteligente pero no técnica, que solo se preocupa por cómo*

esto le ayuda a cerrar tratos. Usa una analogía simple relacionada con el fútbol, idealmente con el equipo Talleres de Córdoba como ejemplo».

Al usar esta estrategia, se mejora la claridad del mensaje y se incrementa la probabilidad de que el receptor entienda el valor práctico de un concepto técnico.

Solicitar opiniones diferentes

Los modelos de lenguaje pueden ser herramientas muy útiles para combatir el pensamiento de grupo (groupthink). En lugar de pedir una única mejor solución, se les puede instruir para que generen múltiples perspectivas —incluso contradictorias— sobre un mismo problema. Esto permite al usuario considerar un espectro más amplio de ideas antes de tomar decisiones.

Prompt (opiniones divergentes): *«Estamos considerando lanzar nuestro producto de limpieza en el mercado de América Latina. Genera un análisis de esta decisión desde tres perspectivas distintas y en conflicto:*

1. **Perspectiva 1:** Un director financiero (CFO), preocupado por el riesgo, los costos de localización y el retorno de inversión a corto plazo.

2. **Perspectiva 2:** un director de marketing (CMO), enfocado en la oportunidad de crecimiento, la adquisición de cuota de mercado y la ventaja competitiva de ser el primero.

3. **Perspectiva 3:** un director de producto (CPO), centrado en los desafíos de adaptar el producto, brindar soporte al cliente en nuevos idiomas y desarrollar la infraestructura local».

Este tipo de enfoque fomenta el pensamiento crítico, ayuda a visualizar riesgos y oportunidades desde diferentes áreas del negocio, y apoya una toma de decisiones más informada y balanceada.

CONTINUAR

Alucinaciones: ¿qué son y cómo manejarlas?

Uno de los desafíos más importantes al trabajar con inteligencia artificial generativa es que, en ocasiones, los modelos proporcionan información incorrecta o incluso completamente inventada con total seguridad, lo que la hace parecer verdadera. A este fenómeno se lo denomina comúnmente «alucinación» en el campo de la inteligencia artificial. El término refleja la idea de que el modelo «ve» —o, más precisamente, genera— contenidos que no están respaldados por la realidad ni por datos verificables de su entrenamiento, de forma análoga a una alucinación humana.

En este sentido, una alucinación es una respuesta del modelo que resulta plausible y detallada, pero que es falsa o carece de fundamento real.

Por ejemplo, el modelo puede afirmar con total convicción una fecha histórica incorrecta, ofrecer una explicación

científica inexistente o incluso inventar referencias bibliográficas. Un caso conocido ocurrió en 2023, cuando un abogado presentó ante un tribunal documentos redactados con ayuda de ChatGPT que citaban hasta seis fallos judiciales como jurisprudencia, los cuales en realidad no existían y habían sido inventados por la inteligencia artificial. El abogado confió en esas citas sin verificarlas y terminó enfrentando sanciones cuando el juez comprobó que eran ficticias (Weiser, 2023). Este episodio ilustra claramente el riesgo que representan las alucinaciones en contextos profesionales.

¿Por qué alucinan los LLM?

Una alucinación no implica «mala fe» por parte del modelo: es una consecuencia directa de cómo aprenden y de cómo solemos evaluarlos. Aunque las alucinaciones han disminuido en versiones más recientes, siguen ocurriendo y es útil comprender su origen para usarlos con criterio.

El origen estadístico durante el preentrenamiento

Aprender «cómo suena» el lenguaje, no qué es verdadero o falso

Durante su entrenamiento inicial, el modelo aprende a predecir la siguiente palabra en enormes cantidades de texto. Este proceso le permite sonar fluido, coherente y «natural». Sin embargo, no se le proporcionan etiquetas que indiquen si cada afirmación es verdadera o falsa. Por eso, puede generar respuestas convincentes... aunque no necesariamente correctas. Es como alguien que escribe bien, pero a quien nunca se le pidió verificar los datos.

Patrones predecibles vs. hechos puntuales.

Los modelos captan muy bien los patrones comunes del lenguaje (gramática, frases típicas, estructuras frecuentes) porque se repiten miles de veces. En cambio, hechos específicos y poco frecuentes —como una fecha exacta o un dato biográfico poco conocido— son más difíciles de capturar, ya que no se derivan de patrones estadísticos generales. Cuando hay lagunas en el contexto o señales débiles, el modelo tiende a rellenar con lo que «suena» adecuado, y ahí pueden aparecer alucinaciones.

La idea de fondo

Podemos imaginar una tarea mental como «¿esto es válido?», que requiere evaluar si una afirmación es correcta. Esta tarea es diferente de simplemente «escribir algo que suene bien». Si un sistema no tiene una forma segura de distinguir entre lo válido y lo

> incorrecto, es natural que ocasionalmente genere frases erróneas que parecen creíbles.

En resumen, los modelos están optimizados para producir lenguaje convincente, no para verificar la veracidad de cada oración. Cuando intervienen hechos raros, ambigüedad o poca información contextual, aumenta la probabilidad de que inventen algo que «encaje» en el estilo del lenguaje, pero que no sea correcto (Kalai et al., 2025).

¿Por qué sobreviven las alucinaciones después del posentrenamiento? El efecto de cómo los evaluamos

Tras el entrenamiento básico, los modelos suelen ajustarse para resultar más útiles y seguros, por ejemplo, para rechazar solicitudes indebidas o adoptar determinados tonos comunicativos. Sin embargo, persiste un factor clave: la forma en que habitualmente evaluamos su rendimiento.

En muchos esquemas de evaluación, la «nota» depende únicamente de acertar. Abstenerse —responder «no lo sé» o «falta información»— puntúa igual que equivocarse. Bajo estas reglas, arriesgar y adivinar puede resultar más

conveniente que ser prudente. Como consecuencia, el modelo aprende de manera indirecta que conviene completar el espacio en blanco antes que reconocer que no dispone de la información necesaria, de forma similar a lo que ocurre en un examen de opción múltiple.

En síntesis, mientras las «reglas del juego» sigan premiando el acierto y castigando por igual el «no lo sé» y el error, existirá una presión estructural a conjeturar en lugar de admitir desconocimiento (Kalai et al., 2025).

¿Cómo identificar una alucinación de IA?

Desarrollar la capacidad para detectar posibles alucinaciones generadas por modelos de lenguaje es una habilidad crítica. Existen ciertas señales de alerta que permiten sospechar de una posible invención por parte del modelo:

DATOS CONCRETOS NO VERIFICABLES	EXCESIVA CONFIANZA SIN MATICES	ESTILO GENÉRICO APLICADO A DATOS SUPUESTAMENTE ESPECÍFICOS	CONOCIMIENTO PREVIO DEL USUARIO

Nombres de estudios, fechas, cifras o referencias que aparentan ser específicas, pero que no se encuentran en una búsqueda rápida o que no concuerdan con información previamente conocida. Si, por ejemplo, ChatGPT menciona una «investigación de la Universidad X de 2022» con gran nivel de detalle, conviene verificar si dicha publicación realmente existe. En muchos casos, el modelo puede inventar títulos de artículos científicos, autores o normativas legales ficticias con el objetivo de respaldar su respuesta.

DATOS CONCRETOS NO VERIFICABLES	EXCESIVA CONFIANZA SIN MATICES	ESTILO GENÉRICO APLICADO A DATOS SUPUESTAMENTE ESPECÍFICOS	CONOCIMIENTO PREVIO DEL USUARIO

Cuando una respuesta se expresa con gran seguridad ante una cuestión ambigua, controversial o poco conocida, es conveniente ser precavidos. Si, ante una pregunta compleja, el modelo responde con una solución definitiva sin proporcionar justificaciones, podría tratarse de una alucinación.

DATOS CONCRETOS NO VERIFICABLES	EXCESIVA CONFIANZA SIN MATICES	ESTILO GENÉRICO APLICADO A DATOS SUPUESTAMENTE ESPECÍFICOS	CONOCIMIENTO PREVIO DEL USUARIO

Una redacción que suena demasiado superficial o generalista puede ser señal de alerta. Por ejemplo, frases como «el Dr. John Doe de Harvard descubrió en 2023 que...» acompañadas de una

explicación ambigua o poco desarrollada pueden indicar que tanto el dato como la fuente han sido inventados.

DATOS CONCRETOS NO VERIFICABLES	EXCESIVA CONFIANZA SIN MATICES	ESTILO GENÉRICO APLICADO A DATOS SUPUESTAMENTE ESPECÍFICOS	CONOCIMIENTO PREVIO DEL USUARIO

Cuando se tiene cierto dominio del tema consultado, resulta más fácil detectar errores o afirmaciones incorrectas. Sin embargo, el riesgo se incrementa en áreas desconocidas para el usuario, donde la redacción convincente del modelo puede inducir a error. Por eso, es recomendable no confiar ciegamente en la respuesta de la IA, especialmente si la información se utilizará para tomar decisiones importantes en contextos laborales o académicos.

En definitiva, la mejor defensa ante posibles alucinaciones es cultivar un escepticismo saludable. La inteligencia artificial no es una fuente de verdad infalible, y mantener una actitud crítica es fundamental para interpretar adecuadamente sus respuestas (Kalai et al., 2025).

Estrategias para mitigar las alucinaciones

Aunque no es posible eliminar completamente las alucinaciones en modelos de lenguaje —una limitación

inherente a la tecnología actual— sí existen diversas estrategias para reducir su frecuencia e impacto.

Formular prompts que fomenten la precisión —

Ciertas técnicas de prompting pueden mejorar la calidad de las respuestas. Pedir razonamiento paso a paso ayuda a exponer el proceso lógico, lo que facilita detectar errores (MIT Sloan Teaching & Learning Technologies, s. f. b.). También es útil solicitar fuentes explícitas: «proporciona tu respuesta junto con la fuente de cada dato importante». Si se utiliza ChatGPT con acceso a internet, esto puede forzar al modelo a traer referencias reales. En contextos profesionales, es preferible utilizar modelos integrados con bases de datos confiables, de modo que el sistema recurra a información validada antes de responder (MIT Sloan Teaching & Learning Technologies, s. f. b.).

Usar herramientas y conocimientos externos —

No se debe depender exclusivamente de la IA. En programación, por ejemplo, siempre es recomendable comprobar el código generado usando un compilador o herramientas de prueba. En entornos educativos, las respuestas pueden tomarse como borradores iniciales, pero es importante contrastarlas con fuentes tradicionales, como libros o consultas docentes. La IA puede ser un acelerador de productividad, pero no debe sustituir a las fuentes confiables en temas críticos.

Verificar datos puntuales

Si la IA entrega un dato central (como una cifra o un porcentaje relevante), conviene validarlo por cuenta propia. Búsquedas rápidas en fuentes como Wikipedia o portales académicos suelen ser suficientes para confirmar o descartar muchas alucinaciones. Esta práctica es una extensión de la alfabetización informacional: al igual que se duda de lo que circula en redes sociales, también hay que aplicar escepticismo ante cualquier dato entregado por un modelo de lenguaje.

Diversificar las consultas

Formular la misma pregunta en diferentes momentos o con distintas palabras puede revelar inconsistencias. Si el modelo da respuestas muy diferentes a la misma consulta, es una señal de que el tema es incierto o problemático. También es útil contrastar la respuesta con otros modelos (como Claude o Gemini); si todos coinciden, es más probable que la información sea confiable.

Modo crítico o validación interna

Una estrategia avanzada consiste en solicitar al modelo que evalúe su propia respuesta. Por ejemplo, luego de una respuesta inicial, se le puede pedir: “Revisa tu respuesta anterior y señala si hay posibles errores o afirmaciones sin fundamento, pensando de forma crítica.” En ocasiones, el modelo reconocerá limitaciones o errores. También se puede usar un segundo prompt de verificación: “¿Qué evidencias respaldan cada afirmación? Si alguna carece de soporte, indícalo.”

> Este tipo de interacción puede promover una salida más cauta, que en estos casos es deseable (MIT Sloan Teaching & Learning Technologies, s. f. b.).

En resumen, nunca se debe utilizar la primera respuesta de la IA tal como está cuando se trata de tareas evaluadas, informes oficiales o decisiones relevantes. La validación externa, el diseño de prompts cuidadosos y el pensamiento crítico son las mejores herramientas para reducir el riesgo de caer en errores producto de una alucinación.

CONTINUAR

Conclusión

A medida que la inteligencia artificial generativa se integra en los ámbitos laboral y académico, saber comunicarse eficazmente con los modelos de lenguaje se perfila como una habilidad clave para los nuevos profesionales. Formular buenos prompts nos permite guiar a la IA para obtener resultados útiles, creativos y pertinentes, potenciando así nuestro desempeño.

Hemos visto cómo técnicas como proporcionar contexto, ser específico, definir el tono, iterar y otras más pueden marcar una gran diferencia en la calidad de las respuestas obtenidas. También abordamos el fenómeno de las alucinaciones de la IA y por qué es vital mantener siempre una mirada crítica y una verificación humana de lo que nos ofrece la máquina.

Es importante recalcar que la IA es una herramienta para ampliar nuestras capacidades, no un reemplazo de ellas. En el mundo laboral actual, quienes sepan colaborar con

sistemas de IA —por ejemplo, utilizando habilidades de prompt engineering— tendrán una ventaja competitiva; sin embargo, esa colaboración implica comprender tanto el potencial como las limitaciones de la tecnología.

En última instancia, detrás de cada interacción humano–IA exitosa hay pensamiento crítico y juicio ético por parte del usuario.

Nota aclaratoria sobre uso de IA

Este material fue asistido con herramientas de IA generativa para tareas de borrador, síntesis, reescritura y apoyo en la organización de contenidos. Cada sección fue revisada, editada y validada por el equipo humano, que verificó la precisión conceptual, la coherencia pedagógica y las fuentes citadas. Se invita a contrastar con las referencias bibliográficas incluidas y la documentación oficial. Dado que los modelos de IA evolucionan con rapidez, ciertas especificaciones técnicas podrían actualizarse; este texto refleja el estado del conocimiento al momento de su elaboración.

CONTINUAR

Referencias

Federiakin, D., Molerov, D., Zlatkin-Troitschanskaia, O., & Maur, A. (2024). *Prompt engineering as a new 21st century skill. Frontiers in Education, 9, Artículo 1366434.* https://doi.org/10.3389/feduc.2024.1366434

Filienko, D., Nizar, M., Roberti, J., Galdamez, D., Jakher, H., Iribarren, S., Yuwen, W., & De Cock, M. (2025). *Transforming Tuberculosis Care: Optimizing Large Language Models For Enhanced Clinician-Patient Communication.* https://doi.org/10.48550/arXiv.2502.21236

Gadesha, V. (s. f.). *What is few-shot prompting? IBM Think.* https://www.ibm.com/think/topics/few-shot-prompting

García-García, J. A. (2025). *Introducción a los prompts – Guía Completa 2024.* https://www.prompt.org.es/docs/prompt-basics/intro

Iyer, R. K. (2025). *Prompt Engineering explained — From basic to advanced techniques. Medium.* https://rkniyer999.medium.com/prompt-engineering-explained-from-basic-to-advanced-techniques-2995198eebb6

Kalai, A. T., Nachum, O., Vempala, S. S., & Zhang, E. (2025). *Why Language Models Hallucinate.* https://arxiv.org/pdf/2509.04664

Kinnoch, D. (s. f.). *Effective Prompting for Planners: Unlocking Large Language Models (LLMs). CoLab Planning.* https://www.colabplanning.co.nz/blog/guide-prompting-llm-ai-urban-planning

Kotha, A., Lee, J., & Zakariasson, E. (2025). *GPT-5 prompting guide. OpenAI Cookbook.* https://cookbook.openai.com/examples/gpt-5/gpt-5_prompting_guide

Kuka, V. (2025). *Few-shot. Learn Prompting.* https://learnprompting.org/docs/basics/few_shot

MIT Sloan Teaching & Learning Technologies. (s. f.). *Effective Prompts for AI: The Essentials. MIT Sloan.* https://mitsloanedtech.mit.edu/ai/basics/effective-prompts/

MIT Sloan Teaching & Learning Technologies. (s. f. b.). *When AI Gets It Wrong: Addressing AI Hallucinations and Bias*. *MIT Sloan*. https://mitsloanedtech.mit.edu/ai/basics/addressing-ai-hallucinations-and-bias/

Schulhoff, S. (2025). *Role prompting*. *Learn Prompting*. https://learnprompting.org/docs/basics/roles

Velásquez-Henao, J. D., Franco-Cardona, C. J., & Cadavid-Higuita, L. (2023). *Prompt Engineering: a methodology for optimizing interactions with Artificial Intelligence language models, like ChatGPT, through prompts in the field of engineering*. *Información tecnológica, 34(5), 181–190*. https://www.redalyc.org/journal/496/49681157002/html/

Weiser, B. (2023). *Here's what happens when your lawyer uses ChatGPT*. *The New York Times*. https://www.nytimes.com/2023/05/27/nyregion/avianca-airline-lawsuit-chatgpt.html

CONTINUAR

Descarga en PDF

FILE

File Attachment Block

No file added