Resuelve tu problema con un guiño

La IA proporcionó las fuentes. Una de ellas hasta existe

Cómo verificar citas y fuentes generadas por herramientas de inteligencia artificial



Max Paradox

La IA proporcionó las fuentes. Una de ellas hasta existe

Cómo verificar citas y fuentes generadas por herramientas de inteligencia artificial

Max Paradox

Rozwiąż swój problem z przymrużeniem oka

ISBN: 9788384906262

ALL-IN Rafał Płonkowski



ÍNDICE

INTRO 5

Capítulo 1 - Bibliografía del país de la fantasía, o por qué la IA inventa fuentes 11

Capítulo 2 - La cita entre comillas, la verdad de vacaciones 21

Capítulo 3 - El enlace funciona, pero ¿adónde lleva? Anatomía de una prueba falsa 31

Capítulo 4 - «Pero si la IA ya lo comprobó», o las excusas más caras del mundo 41

Capítulo 5 - No busques una aguja en una biblioteca, o cómo encontrar una fuente sin perder la cabeza 53

Capítulo 6 - Cuando la respuesta encaja demasiado bien, o la trampa de nuestras propias expectativas 64

Capítulo 7 - El teléfono escacharrado con traje y corbata, o cómo se pierden las fuentes entre personas y documentos 75

Capítulo 8 - El PDF lo aguanta todo, o cuando la IA interpreta mal documentos auténticos 86

Capítulo 9 - Cinco minutos para la verdad, o cómo realizar una primera auditoría de una respuesta de IA 98

Capítulo 10 - De la cita al original, o una investigación con un final auténtico 109

Capítulo 11 - Los números no mienten, pero sus descripciones tienen una imaginación desbordante 120

Capítulo 12 - Construye una bibliografía de la que no tengas que avergonzarte 131

Capítulo 13 - Enseña a la IA a mostrar pruebas, no a causar buena impresión 142

Capítulo 14 - De una respuesta de IA a una decisión que puedas defender 153

Capítulo 15 - Dos fuentes, tres verdades y una persona al borde de la desesperación 165

Capítulo 16 - La fuente que empieza a dar órdenes, o por qué no debes obedecer todas las instrucciones de un documento 176

Capítulo 17 - Cuando no puedes comprobarlo todo, o el arte de aceptar honestamente lo que queda pendiente 188

Capítulo 18 - Una verificación que no termina al pulsar «Enviar» 199

Capítulo 19 - Tu propio radar para detectar disparates, o cómo evitar volver a los viejos hábitos 211

Capítulo 20 - La IA puede escribir con seguridad. Tú puedes saber en qué se basa 223

INTRO

Imagina que estás preparando un informe importante. La fecha límite se acerca a la velocidad de un repartidor que, por supuesto, llega justo cuando acabas de meterte en la ducha. Tienes los datos, tienes los argumentos y solo te faltan unas cuantas fuentes sólidas. Abres una herramienta de inteligencia artificial y le pides publicaciones que respalden una afirmación determinada. Al cabo de unos instantes, recibes cinco referencias bibliográficas impecables. Los autores tienen nombre y apellido, los artículos tienen título, las revistas tienen números de edición y algunas publicaciones incluso incluyen identificadores DOI. Todo parece tan profesional que te entran ganas de imprimir la respuesta, encuadernarla en piel y colocarla en una estantería junto a la enciclopedia.

Solo hay un pequeño problema. El primer artículo no existe. El segundo existe, pero trata de otra cosa. El tercero tiene un autor real y un título inventado. El cuarto conduce a una página que dejó de funcionar hace mucho tiempo. El quinto existe de verdad, pero sus autores sostienen prácticamente lo contrario de lo que ha escrito la IA.

Enhorabuena. Acabas de recibir una bibliografía en la que la realidad hace una aparición especial.

Sin embargo, lo más inquietante no es que la inteligencia artificial sea capaz de inventarse una fuente. Lo verdaderamente fascinante es la convicción con la que puede presentarla. No aparece ninguna advertencia que diga: «Atención: los siguientes nombres podrían ser fruto de la imaginación literaria del procesador». No escuchas una tos nerviosa. No ves ningún rubor digital. Recibes una

respuesta serena y ordenada, que parece redactada por alguien que acaba de regresar de la biblioteca después de pasar tres días estudiando publicaciones científicas.

La realidad es que un modelo de lenguaje puede producir un texto que se parezca a una bibliografía correcta sin disponer de fundamentos suficientes para confirmar la veracidad de cada uno de sus elementos. Y esa diferencia es fundamental. Saber redactar una referencia bibliográfica convincente no equivale a localizar una publicación, leerla y comprobar si realmente respalda la afirmación presentada. Del mismo modo, saber dibujar un mapa del tesoro no significa que bajo la palmera señalada haya un cofre lleno de oro. A veces lo único que hay es la factura de las vacaciones.

Y, sin embargo, utilizar la IA resulta realmente cómodo. Ayuda a organizar ideas con rapidez, comprender cuestiones complejas, preparar borradores y encontrar posibles direcciones para seguir investigando. El problema comienza cuando confundimos la comodidad con la evidencia. Una respuesta suena convincente y dejamos de preguntarnos de dónde procede la información. Recibimos un enlace y suponemos que conduce a contenido fiable. Vemos unas comillas y creemos que las palabras citadas fueron pronunciadas realmente por alguien.

Así es como podemos terminar publicando una cita que nadie dijo jamás, mencionando una investigación inexistente o transmitiendo información errónea a nuestros compañeros de trabajo acompañada de la expresión «según los expertos». Los expertos descubren después que participaron en un estudio del que ni siquiera habían oído hablar. Probablemente se sorprendan tanto como un

unicornio al enterarse de que alguien lo ha inscrito en el registro de animales de granja.

Este libro no pretende convencerte de que la inteligencia artificial es inútil. Eso sería tan razonable como tirar una calculadora porque alguien introdujo números equivocados. Se trata de aprender a distinguir entre recibir ayuda para buscar información y confirmar que esa información es verdadera. Puedes utilizar la IA intensivamente, ahorrar tiempo y seguir manteniendo el control sobre aquello que finalmente aceptas como un hecho.

Para empezar, conviene comprender una verdad algo incómoda: no todas las fuentes incorrectas son completamente inventadas. Si cada referencia falsa condujera a un artículo titulado «Investigación Extremadamente Importante que Acabo de Inventarme», la vida sería mucho más sencilla. Sin embargo, lo habitual es encontrarse con errores más sutiles. Una publicación auténtica se atribuye al autor equivocado. Un informe real aparece con un año de publicación incorrecto. Una paráfrasis se presenta como si fuera una cita literal. Un enlace conduce a una página auténtica, pero su contenido no respalda la conclusión que se está defendiendo.

Por eso, limitarse a preguntar «¿Existe esta fuente?» no es suficiente. También hay que saber si se ha identificado el documento correcto, si se reproduce fielmente su contenido, si el autor de la fuente realmente aborda la cuestión y si la conclusión extraída no va más allá de las pruebas disponibles. Son varias comprobaciones diferentes, aunque cuando tenemos prisa resulta tentador meterlas todas en el mismo cajón con una etiqueta que diga «Verificado». El cajón parece perfectamente ordenado. Su contenido ya es otra historia.

En la práctica, trataremos las respuestas de la IA como un punto de partida para la verificación, no como un certificado automático de veracidad. Aprenderás a reconocer señales de alerta en las bibliografías, localizar documentos originales, comprobar autores y fechas, verificar identificadores de publicaciones y encontrar las citas dentro de su contexto real. También descubrirás por qué diferencias aparentemente pequeñas entre «el estudio demostró», «los autores sugieren» y «se observó en determinadas condiciones» pueden cambiar por completo el significado de una información.

Esto no significa que tengas que convertirte en detective bibliográfico profesional. No necesitas un despacho forrado de libros, una lupa ni música dramática cada vez que abres una nueva pestaña del navegador. Lo que necesitas es un método que te permita determinar qué debes comprobar primero, dónde buscar confirmación y cuándo las pruebas disponibles son insuficientes. De lo contrario, acabarás dedicando tanto tiempo a verificar el título de un artículo poco relevante como a comprobar una información de la que depende una decisión importante.

Al fin y al cabo, no todas las situaciones exigen el mismo nivel de precisión. Si estás preparando una nota informal para uso personal, puedes aceptar un mayor grado de incertidumbre, siempre que no lo ocultes. Si estás elaborando una publicación científica, material educativo, un documento profesional o un texto relacionado con la salud, el derecho o las finanzas, las consecuencias de un error pueden ser mucho más graves. En esos casos, verificar las fuentes no es una precaución exagerada. Es parte de la responsabilidad que implica transmitir información.

También será importante aprender a reconocer cuándo no has conseguido confirmar algo. Suena sencillo, pero las herramientas actuales nos han acostumbrado a obtener respuestas inmediatas. Un espacio vacío en una bibliografía parece casi un fracaso personal. Sin embargo, afirmar con sinceridad «No he encontrado una confirmación fiable de esta cita» tiene mucho más valor que presentar una referencia perfectamente formateada que no conduce a ninguna parte. La ausencia de pruebas no siempre demuestra que algo sea falso, pero desde luego tampoco autoriza a presentarlo como verificado.

Con el tiempo empezarás a darte cuenta de que la habilidad más importante no consiste en pillar a la IA cometiendo errores. Consiste en formular las preguntas adecuadas y distinguir entre una respuesta y una prueba. Un modelo puede ayudarte a encontrar pistas, señalar posibles publicaciones, proponer términos de búsqueda y organizar resultados. Sin embargo, la confirmación definitiva exige consultar los materiales correspondientes, preferiblemente en la página del editor original o en otra fuente fiable que permita acceder al documento.

Así que, si alguna vez has copiado una referencia bibliográfica sin abrir la fuente, has creído una cita simplemente porque sonaba inteligente o has considerado que la presencia de un enlace era una prueba suficiente, estás en el lugar adecuado. No vamos a organizar un juicio público contra tu navegador. Aprenderemos algo mucho más útil: aprovechar las posibilidades de la inteligencia artificial sin concederle el derecho a decidir qué es verdad.

Porque la inteligencia artificial puede preparar una lista de diez fuentes en cuestión de segundos. Tu tarea no consiste en admirar lo larga que es la lista. Tu tarea consiste en saber

cuántas de esas diez referencias merecen realmente aparecer en tu texto.

Y si resulta que solo una existe, al menos sabrás cuál es.

Capítulo 1 - Bibliografía del país de la fantasía, o por qué la IA inventa fuentes

Lunes, 9:17 de la mañana. En una sala de reuniones estás presentando los resultados de un análisis que preparaste la noche anterior. Todo marcha estupendamente. Los gráficos son claros, las conclusiones son razonables y la presentación incluso incluye una diapositiva con las fuentes, algo que en el mundo de las presentaciones corporativas a veces se considera casi una muestra de educación aristocrática. En determinado momento, alguien pregunta de dónde procede la información sobre la eficacia de la solución que estás exponiendo. Sin dudarlo, señalas una publicación científica que la inteligencia artificial te recomendó el día anterior.

Tu interlocutor introduce el título del artículo en un buscador. Nada. Comprueba el nombre del autor. Encuentra a un investigador, pero trabaja en un campo completamente diferente. Intenta localizar la revista. Esa sí existe, aunque el número indicado no contiene ninguna publicación semejante. La sala se queda en silencio. Incluso el proyector, que hasta entonces emitía sonidos parecidos al despegue de una avioneta, parece funcionar con mayor discreción.

Acabas de descubrir que tu bibliografía tiene más imaginación que el departamento creativo.

Es una situación especialmente desagradable porque el error no tenía aspecto de error. No había faltas de ortografía, palabras aleatorias ni frases sospechosas. La fuente tenía autor, fecha, título, nombre de la revista y número de volumen. Todos los elementos estaban colocados donde

correspondía. Solo faltaba un pequeño detalle: una publicación real que alguien pudiera consultar.

Conviene comprender por qué se producen estas situaciones. De lo contrario, resulta fácil adoptar una de dos posturas extremas. La primera sostiene que la inteligencia artificial es un bibliotecario genial y que debemos confiar en ella. La segunda la presenta como un estafador digital que intenta dejarnos en ridículo deliberadamente. Ambas visiones son llamativas, pero ninguna ayuda a utilizar correctamente la herramienta.

Un modelo de lenguaje no necesita tener una intención real de engañarte para generar información falsa. Puede limitarse a crear una respuesta que encaje con los patrones aprendidos durante su entrenamiento, aunque no esté suficientemente vinculada a un documento auténtico. Esa es la diferencia fundamental entre la apariencia correcta de una respuesta y la veracidad de su contenido.

Cuando una referencia elegante se hace pasar por una prueba

Imagina una tarea que consiste en completar la frase: «Según una investigación publicada en la revista...». Una persona que conozca el idioma y las normas básicas de la escritura científica puede proponer una continuación natural. Sin embargo, eso no significa que haya consultado el archivo de la editorial correspondiente. Del mismo modo, un modelo puede generar referencias bibliográficas convincentes utilizando relaciones reconocidas entre nombres, títulos, disciplinas y formas habituales de citar publicaciones.

Dependiendo de su diseño y de las funciones disponibles, la herramienta también puede buscar en internet, leer

documentos o consultar bases de datos externas. Estas posibilidades mejoran las condiciones para verificar la información, pero no garantizan por sí solas que la respuesta final sea correcta. El modelo puede interpretar mal el material encontrado, combinar información procedente de dos publicaciones o atribuir a una fuente auténtica una conclusión que sus autores nunca formularon.

Por tanto, el problema tiene varios niveles. A veces la fuente no existe. Otras veces existe, pero está descrita incorrectamente. En otras ocasiones, todos los datos bibliográficos son correctos y el error se encuentra en la interpretación del texto. Esta última situación resulta especialmente traicionera porque una simple búsqueda para comprobar si la publicación aparece en internet no permite descubrir el fallo.

Pongamos un ejemplo ficticio. Pides a la IA investigaciones sobre la influencia de las pausas breves en la concentración durante el trabajo de oficina. Recibes el título «The Impact of Microbreaks on Cognitive Performance in Office Environments», los nombres de los autores, el año de publicación y el nombre de una revista de psicología. Todo suena extraordinariamente plausible. El tema encaja con la disciplina, el título transmite seriedad académica y los nombres parecen pertenecer a personas que participan regularmente en congresos.

Eso no significa que el artículo indicado exista. El título de este ejemplo es una ilustración, no una referencia bibliográfica confirmada. Así es precisamente como debes tratar cualquier propuesta de fuente que no haya sido verificada: como información pendiente de comprobación, aunque tenga una apariencia absolutamente profesional.

Curiosamente, cuanto más detallada es una respuesta, mayor puede ser la sensación de credibilidad que provoca. Los números de página, las fechas de publicación y los identificadores DOI recuerdan a los sellos oficiales. Los vemos y asumimos inconscientemente que alguien ya ha realizado las comprobaciones necesarias. Sin embargo, una respuesta detallada puede ser simplemente eso: detallada. Un error extraordinariamente preciso sigue siendo un error, solo que mejor vestido.

Ni siquiera hace falta inventarse una revista para meterse en problemas. Basta con combinar el nombre auténtico de una editorial prestigiosa con un artículo inexistente. El elemento verdadero proporciona al conjunto una apariencia de solidez. Es como si alguien te diera la dirección ficticia de un apartamento situado en un rascacielos que realmente existe. El edificio está ahí, la calle es correcta, pero detrás de la puerta indicada solo hay una pared.

Por qué nos dejamos convencer tan fácilmente

Nuestros problemas no proceden únicamente del funcionamiento de la tecnología. La otra mitad del asunto está relacionada con la manera en que evaluamos la información. Cuando un texto es fluido, está bien organizado y utiliza un tono seguro, tendemos a considerarlo más competente. Especialmente cuando trata un tema que nosotros mismos no dominamos suficientemente.

Tampoco ayuda que normalmente pidamos apoyo a la IA precisamente cuando queremos ahorrar tiempo. Necesitamos material con rapidez, de modo que la bibliografía recibida encaja perfectamente con nuestras expectativas. Verificarla supone trabajo adicional. Abrir publicaciones, comprobar autores y comparar citas suena bastante menos atractivo que pulsar el botón «Copiar».

A esto se suma una tentación más sutil: buscar confirmación de nuestras propias opiniones. Si la herramienta presenta cinco fuentes que respaldan nuestra tesis, sentimos satisfacción. Si señalara cinco publicaciones que cuestionan esa misma tesis, probablemente empezaríamos a formular preguntas adicionales. Precisamente por eso debemos prestar especial atención a las fuentes que dicen exactamente lo que queríamos escuchar.

Por ejemplo, estás preparando argumentos a favor de implantar una determinada forma de organizar el trabajo. Preguntas: «¿Qué investigaciones demuestran que este método aumenta la productividad?». La propia pregunta presupone que existen pruebas que respaldan esa afirmación. El modelo puede responder siguiendo la expectativa expresada en lugar de cuestionar su fundamento.

Sería más razonable preguntar: «¿Qué publicaciones fiables investigan la eficacia de este método? Indica tanto los resultados favorables como los desfavorables y señala claramente cualquier información que no puedas confirmar». Esta formulación no garantiza una respuesta libre de errores, pero cambia la dirección de la búsqueda. Dejas de encargar pruebas para una conclusión que ya habías preparado de antemano.

Esto también importa cuando utilizas una herramienta que realmente tiene acceso a un buscador. Internet contiene una enorme cantidad de materiales de calidad muy desigual. Encontrar una página no significa que la afirmación que aparece en ella sea verdadera. Buscar información y evaluar su credibilidad son dos actividades distintas.

Las cuatro caras de una fuente falsa

En la práctica conviene reconocer cuatro situaciones fundamentales, porque cada una exige una reacción diferente.

La primera es la ficción bibliográfica absoluta. La publicación, los autores o la revista no corresponden a ningún documento real. Puedes pasar media hora buscando un artículo que únicamente existe dentro de una respuesta generada. Si después de consultar razonablemente catálogos y bases de datos no encuentras confirmación, no lo incluyas en tu bibliografía como si fuera auténtico. No necesitas demostrar que el documento jamás existió en toda la historia de la humanidad. Basta con reconocer que no has conseguido verificarlo.

La segunda situación consiste en una mezcla de datos. El artículo existe, pero se indica un año, un título o unos autores incorrectos. En ocasiones, el modelo construye una referencia bibliográfica combinando fragmentos de varias publicaciones auténticas. Para detectar este error hay que comparar la descripción completa con el registro bibliográfico original. No conviene corregir los elementos a ojo, porque podrías acabar creando accidentalmente otro híbrido.

La tercera forma es la cita distorsionada. Un autor escribió realmente una frase determinada, pero la IA modifica algunas palabras, elimina una advertencia importante o combina dos fragmentos en una única declaración. El significado puede seguir siendo parecido, pero el texto colocado entre comillas debe corresponder al original. Si quieres expresar una idea con tus propias palabras, utiliza una paráfrasis e identifica adecuadamente la fuente. No disfraces una paráfrasis de cita literal.

La cuarta situación es la más astuta: la fuente existe, la cita es correcta, pero no respalda la afirmación presentada. Una publicación describe una relación entre dos variables y la IA la presenta como una prueba de causalidad. Un estudio examina un grupo pequeño y específico, pero la respuesta formula una conclusión aplicable a toda la población. Los autores hablan de una posibilidad y el resumen generado anuncia una certeza absoluta.

Por tanto, puedes tener un artículo auténtico, un enlace correcto y, aun así, un argumento equivocado. La bibliografía queda absuelta, pero la conclusión acaba de huir de la escena del crimen.

El primer filtro antes de empezar a buscar

No necesitas realizar inmediatamente un análisis exhaustivo de cada publicación. En el trabajo cotidiano resulta útil disponer de un filtro inicial que permita decidir qué referencias requieren una comprobación más detallada. Empieza leyendo la respuesta de la IA como una propuesta y no como una investigación terminada.

Presta atención a las fuentes presentadas sin datos completos, a las que parecen encajar sospechosamente bien con tu pregunta y a las que contienen afirmaciones muy categóricas sin señalar fragmentos concretos del documento. Estas características no demuestran que sean falsas, pero justifican una comprobación. Del mismo modo, la ausencia de autor o fecha no siempre significa falta de credibilidad, porque no todos los materiales valiosos tienen la estructura de una publicación científica.

Para cada referencia importante, intenta responder a tres preguntas. ¿Puedo encontrar el documento independientemente de la respuesta de la IA? ¿Coincide la

descripción bibliográfica con el documento real? ¿Sé qué fragmento concreto respalda mi afirmación? Si te quedas atascado en la primera pregunta, no saltes automáticamente a la tercera fingiendo que todo está en orden.

Supongamos que has recibido cinco fuentes para una presentación. Encuentras rápidamente dos en catálogos fiables. Una tiene el año equivocado, pero puedes identificar el documento original. Las otras dos resultan imposibles de localizar. Eso no significa que tengas que tirar todo el trabajo. Puedes conservar las referencias confirmadas, corregir los datos de la tercera y eliminar las dos que no has podido verificar. Lo importante es no tratar las cinco como si fueran igualmente fiables.

Si una fuente es fundamental para tomar una decisión importante, confirmar que existe será apenas el comienzo. También será necesario evaluar el contenido y las limitaciones de la publicación. En cambio, si se trata de material complementario, quizá puedas prescindir de una referencia dudosa en lugar de dedicar horas a buscarla.

También merece la pena conservar un registro de tus comprobaciones. No tiene por qué ser un sistema complicado. Una nota breve como «confirmado en la página de la editorial», «no se ha encontrado el texto completo» o «la fuente no respalda la conclusión» suele ser suficiente para recordar, una semana después, por qué utilizaste o descartaste una publicación determinada.

La IA como ayudante, no como testigo principal

El cambio más útil consiste en separar dos funciones. La inteligencia artificial puede resultar excelente para preparar preguntas, organizar conceptos y diseñar estrategias de búsqueda. Sin embargo, no debería convertirse

automáticamente en el testigo definitivo de la veracidad de la información que ella misma ha generado.

Cuando preguntas: «¿Son auténticas las fuentes que me has proporcionado?», puedes recibir otra respuesta convincente que también necesite verificación. Eso no constituye una comprobación independiente. La herramienta puede corregir un error, pero también puede mantener una imprecisión anterior. La simple afirmación «lo he comprobado» no sustituye la posibilidad de señalar qué se ha verificado realmente y dónde.

Es preferible utilizar la IA para preparar una lista de elementos que requieren confirmación y después compararlos con los documentos originales. También puedes pedirle que indique las limitaciones de su propia respuesta. De este modo recibirás ayuda para organizar el trabajo sin trasladar toda la responsabilidad al sistema.

No hace falta tratar cada frase generada como si fuera un sospechoso de una investigación criminal. Basta con adaptar el nivel de atención a las consecuencias de un posible error. Una información utilizada como inspiración informal tiene un peso diferente al de una cita incluida en un trabajo universitario o una fuente que sirve de base para una decisión empresarial. Precisamente en esos casos importantes necesitas algo más que una referencia bibliográfica elegante.

Así que, la próxima vez que recibas una bibliografía que parezca el resultado de un mes de trabajo de todo un equipo de investigadores, no la descartes automáticamente. Pero tampoco la aceptes sin comprobarla. Trátala como una lista de posibles direcciones que merece la pena investigar.

Porque incluso la bibliografía mejor presentada del mundo seguirá siendo inútil si la mitad de sus autores viven exclusivamente en la imaginación del modelo.

Capítulo 2 - La cita entre comillas, la verdad de vacaciones

Las comillas son uno de los signos tipográficos más pequeños a los que concedemos un poder sorprendentemente grande. Basta con colocar unas palabras entre ellas, añadir el nombre de un personaje famoso y, de repente, una frase corriente empieza a parecer una declaración histórica. Si al lado aparece una fecha, el título de un libro o el nombre de una conferencia, la credibilidad aumenta todavía más. Dejamos de preguntarnos si alguien pronunció realmente esas palabras y empezamos a pensar cuál sería la mejor forma de utilizarlas en nuestra presentación.

Un día estás preparando un artículo sobre la toma de decisiones. Pides a la IA una cita breve de un pensador conocido que destaque la importancia del pensamiento crítico. Recibes una frase tan acertada que te entran ganas de colocarla en un póster encima del escritorio. Tiene un ritmo adecuadamente solemne, el autor es reconocible y el mensaje encaja perfectamente con tu texto.

Sin embargo, cuando intentas localizar la cita en la publicación original, aparece un problema. El buscador muestra cientos de páginas que repiten la misma frase, pero ninguna señala el lugar concreto donde apareció por primera vez. Un portal menciona un libro, otro un discurso y un tercero proporciona una fecha que no puede relacionarse con ningún acontecimiento confirmado. Un cuarto cita al primero, un quinto al cuarto y un sexto asegura que la frase es «de conocimiento general».

Acaba de formarse una orquesta digital en la que todos los músicos interpretan la misma partitura equivocada.

El problema de las citas falsas es anterior a la inteligencia artificial. Desde hace mucho tiempo atribuimos a personajes conocidos declaraciones que suenan suficientemente inteligentes, divertidas o inspiradoras. Internet ha facilitado su difusión y las herramientas de IA pueden acelerar todavía más la creación de nuevas versiones. Sin embargo, no todas las atribuciones incorrectas fueron originadas por un modelo de lenguaje. Algunos errores ya estaban presentes en los materiales disponibles antes de que el modelo fuera entrenado.

Precisamente por eso, verificar citas exige mayor precisión que limitarse a preguntar quién es su autor. Hay que determinar qué palabras se pronunciaron realmente, en qué contexto, dónde quedaron registradas y si la persona indicada es efectivamente su autora. Solo entonces las comillas empiezan a cumplir su verdadera función.

Cuando una paráfrasis se pone el traje de otra persona

El problema más habitual no tiene por qué consistir en inventar completamente una declaración. A veces la IA reconoce el sentido general de una idea y la reformula con sus propias palabras, para después presentarla como una cita literal. El resultado es una frase fiel a un determinado planteamiento, pero infiel a las palabras reales del texto.

Supongamos que el autor de una publicación escribe: «En el grupo de participantes analizado, las pausas regulares se asociaron con un menor cansancio declarado». La IA puede transformar este fragmento en una afirmación mucho más contundente: «Las pausas regulares reducen el cansancio

de los trabajadores». A primera vista, ambas frases tratan un asunto parecido. Sin embargo, la diferencia es enorme.

La primera describe una relación observada en un grupo determinado. La segunda sugiere una relación causal general. Se ha eliminado la limitación relativa a las personas estudiadas, ha desaparecido la información de que el cansancio era declarado por los participantes y una formulación prudente ha sido sustituida por una afirmación categórica.

Si colocas la segunda frase entre comillas y la firmas con el nombre del autor, le estarás atribuyendo palabras que, en este ejemplo, nunca pronunció. Además, habrás cambiado el significado de su declaración. Eso ya no es un inocente retoque lingüístico.

Recuerda que una cita literal y una paráfrasis cumplen funciones diferentes. La cita sirve para reproducir fielmente unas palabras concretas. La paráfrasis permite expresar una idea ajena utilizando tu propio lenguaje, conservando su significado e identificando adecuadamente la fuente. En ambos casos necesitas comprender el original, pero solo en el primero se exige correspondencia con las palabras exactas.

La inteligencia artificial puede ayudarte a preparar una paráfrasis, siempre que tengas acceso al texto original y puedas evaluar si es correcta. En cambio, cuando solicitas una cita exacta, necesitas un fragmento real del documento. El modelo no debería sustituir las palabras que faltan con su propia creatividad, aunque lo haga con una elegancia literaria admirable.

Las citas traducidas añaden otra dificultad. Una traducción al polaco de una declaración inglesa no

reproduce literalmente las palabras originales en el mismo idioma. Puede citarse como traducción, pero es necesario indicar adecuadamente su carácter. Si utilizas una traducción publicada, conviene identificar su fuente. Si la realizas por tu cuenta, no presentes tu versión como si fuera la única formulación autorizada.

Esta diferencia es especialmente importante en textos científicos y profesionales, donde una sola palabra puede modificar el alcance de una afirmación. El término inglés «may» no equivale a «must», y «associated with» no significa automáticamente «caused by». Al pulir el lenguaje, podemos transformar inadvertidamente una conclusión prudente en una garantía categórica.

Setecientas páginas repiten lo mismo. ¿Quién lo ha comprobado?

Una de las características más engañosas de internet es la posibilidad de reproducir una información muchas veces sin aumentar el número de pruebas independientes. Una cita puede aparecer en cientos de páginas, miles de publicaciones y decenas de presentaciones, aunque todos esos materiales procedan, en última instancia, de una única entrada que nadie verificó.

Imagina que alguien publica una frase impactante atribuida a un científico famoso. Otro autor la copia en un artículo, una tercera persona la coloca en una imagen y alguien más la incorpora a una colección de citas inspiradoras en internet. Después de varios años, el buscador devuelve una enorme cantidad de resultados. La simple repetición hace que la información empiece a resultarnos familiar, y esa familiaridad puede confundirse con la verdad.