Evita que te engañen gracias al prebunking

Descripción

Introducción

Un mensaje llega a uno de tus grupos familiares de WhatsApp. Advierte de un supuesto peligro que “los medios no quieren contar”, cita a un médico que nadie conoce y termina con una petición urgente de “Compártelo antes de que lo borren”.

Todavía no sabemos si la información es verdadera y, sin embargo, el mensaje ya ha empezado a influir en nosotros. Ha despertado miedo. Insinúa la existencia de un enemigo, utiliza la autoridad de un supuesto experto y ha introducido una urgencia artificial que dificulta detenerse a comprobar los datos. Su eficacia no depende únicamente de que aceptemos lo que afirma, también depende de que reaccionemos antes de pensarlo.

Algo parecido ocurre cuando una oferta comercial asegura que solo estará disponible durante unos minutos, cuando un vídeo presenta dos únicas soluciones para un problema complejo, cuando una publicación utiliza una cifra espectacular sin explicar de dónde procede o cuando alguien intenta neutralizar cualquier crítica futura afirmando que vendrá de personas interesadas en ocultar la verdad.

Corregir estos mensajes después de que hayan sido aceptados es posible, pero no siempre resulta sencillo. La información falsa puede continuar influyendo incluso cuando sabemos que ha sido desmentida. Además, las correcciones suelen llegar tarde, después de que el contenido haya circulado, activado emociones, reforzado identidades y

condicionado decisiones.

El prebunking propone intervenir antes. También conocido como inoculación psicológica, consiste en advertir anticipadamente de una posible manipulación, mostrar una versión debilitada de la estrategia utilizada y enseñar a reconocerla. No pretende decirnos qué debemos pensar sobre cada tema, sino ayudarnos a detectar cuándo alguien intenta alterar las condiciones bajo las que pensamos.

La pregunta que guía este artículo es sencilla: ¿Podemos prepararnos para reconocer una manipulación antes de caer en ella?

La evidencia científica indica que sí, aunque con importantes matices. El prebunking puede mejorar nuestra capacidad para distinguir contenidos fiables de contenidos engañosos, pero no nos convierte en personas inmunes a la desinformación. Sus efectos suelen ser modestos, pueden debilitarse con el tiempo y no siempre se traducen en cambios reales de conducta.

Aun así, aprender cómo funciona ofrece algo valioso: una pequeña distancia entre el estímulo y la respuesta. Un espacio en el que podemos detenernos, recuperar el contexto y decidir con mayor autonomía.

La extraña idea de vacunar una creencia

La teoría de la inoculación psicológica fue desarrollada durante la década de 1960 por el psicólogo social William J. McGuire. Su propuesta partía de una analogía médica. Del mismo modo que una vacuna expone al organismo a una versión controlada de una amenaza para preparar sus defensas, una persona podría exponerse a versiones debilitadas de argumentos persuasivos para desarrollar resistencia frente a intentos posteriores de influencia.

La comparación es muy interesante, pero conviene utilizarla con prudencia. Una creencia no es un virus, la mente no genera anticuerpos contra la manipulación y una intervención psicológica no produce inmunidad permanente. El prebunking modifica probabilidades, pero no instala una protección infalible.

La formulación clásica de la inoculación incluye dos componentes fundamentales.

El primer paso consiste en comunicar que alguien podría intentar influir sobre una creencia, una decisión o nuestra autonomía.

Por ejemplo, el mensaje “Es posible que recibas mensajes que intenten presionarte para que tomes una decisión sin comprobar la información” es una advertencia que genera una sensación moderada de amenaza. No necesariamente miedo, sino la percepción de que alguien podría estar intentando dirigir nuestra respuesta. Esa percepción puede aumentar la atención y motivarnos para examinar mejor el mensaje.

La advertencia no debería limitarse a afirmar que “todo es mentira”. Su función es señalar el mecanismo de influencia. Podrían intentar activarte emocionalmente, quizá presenten un falso dilema, pueden utilizar una autoridad aparente, tal vez seleccionen únicamente los datos favorables o podrían presionarte para que actúes de inmediato.

El segundo componente consiste en presentar un ejemplo sencillo y explicar cómo funciona.

Imaginemos esta frase “Compártelo ahora, antes de que lo censuren”. Una intervención de prebunking podría explicar que esta expresión combina dos tácticas. Por un lado, crea urgencia para reducir el tiempo disponible para verificar. Por otro, convierte una posible desaparición futura del contenido en una supuesta prueba de censura.

La persona aprende así una respuesta defensiva. La urgencia no demuestra que la información sea cierta. Cuanto mayor sea la presión para actuar, mayor debería ser la necesidad de detenerse.

Lo importante no es memorizar esta frase, sino comprender el patrón. La próxima vez, la estrategia puede adoptar otra forma tipo “Solo quedan tres unidades”, “Tienes que decidir ahora”, “Difúndelo antes de que lo eliminen”, “No tendrás otra oportunidad” o “Si no actúas hoy, será demasiado tarde”.

Los contenidos son diferentes, pero la arquitectura psicológica es similar. Ante esto, la investigación contemporánea suele diferenciar entre dos modalidades diferentes para inocular contra afirmaciones o contra técnicas

La inoculación basada en contenidos prepara frente a una afirmación concreta. Por ejemplo, anticipar un rumor determinado sobre una vacuna, una elección o una tecnología y proporcionar información para refutarlo.

La inoculación basada en técnicas enseña a reconocer mecanismos generales de manipulación, como el lenguaje emocional, los falsos dilemas, la suplantación de expertos, la descontextualización o el uso selectivo de datos.

Esta segunda modalidad resulta especialmente interesante porque puede transferirse entre situaciones distintas. Quien aprende a identificar una falsa dicotomía en un discurso político puede reconocer la misma estructura en una campaña comercial, una discusión laboral o una publicación sobre salud.

Diversos experimentos con vídeos breves y juegos interactivos han enseñado a detectar tácticas como la apelación emocional, las falsas dicotomías, la incoherencia, los ataques personales y la búsqueda de chivos expiatorios. En conjunto, los participantes mejoraron su capacidad para reconocer mensajes manipuladores.

La idea es que no podemos aprender una respuesta para cada posible engaño, pero sí podemos familiarizarnos con un repertorio relativamente limitado de técnicas que se repiten en miles de mensajes diferentes.

.

Qué funciona realmente y qué estamos exagerando

El prebunking se ha convertido en una intervención atractiva porque parece ofrecer una respuesta escalable frente a la desinformación. Un vídeo, un juego o una campaña educativa pueden llegar a muchas personas antes de que aparezca un contenido engañoso.

Pero que una idea sea atractiva no significa que su eficacia esté garantizada. La cuestión relevante es qué muestran conjuntamente los estudios. Mejora el discernimiento, pero no hace milagros

Una revisión sistemática y metaanálisis publicada en 2023 reunió 42 estudios independientes y más de 42.000 participantes. Sus resultados indicaron que la inoculación psicológica reducía la credibilidad atribuida a la desinformación y mejoraba la capacidad para diferenciar información verdadera de información falsa. Los efectos fueron, en general, pequeños o moderados.

Esto importa por dos motivos. El primero es que existe una base empírica razonable para considerar el prebunking una intervención útil. No se trata únicamente de una metáfora convincente. El segundo es que sus efectos no justifican hablar de inmunidad mental. Una mejora media en un experimento no significa que todas las personas reconocerán todas las manipulaciones ni que lo harán cuando estén cansadas, asustadas, enfadadas o presionadas por su grupo.

Un metaanálisis posterior utilizó la teoría de detección de señales para diferenciar dos posibilidades. Una intervención podía mejorar realmente la discriminación entre noticias fiables y no fiables o podía limitarse a hacer que la gente desconfiara de todo. Los resultados favorecieron la primera interpretación. Las intervenciones mediante vídeos y juegos mejoraban el discernimiento sin producir necesariamente un aumento general de la incredulidad.

Este matiz es fundamental. El pensamiento crítico no consiste en negar sistemáticamente la información, sino en ajustar el nivel de confianza a la calidad de la evidencia.

Pero uno de los resultados más relevantes del metaanálisis de 2023 fue que la inoculación mejoraba diferentes medidas de discernimiento, pero no reducía de forma clara la intención general de compartir desinformación.

A primera vista puede parecer contradictorio. Sin embargo, las personas no comparten contenidos únicamente porque los consideren ciertos.

También pueden difundirlos porque expresan su identidad, atacan al grupo contrario, producen indignación, resultan divertidos, refuerzan la pertenencia, generan conversación o, simplemente, confirman algo que desean que sea verdad.

Una persona puede sospechar que un titular está exagerado y compartirlo de todos modos porque representa lo que siente. También puede difundir una historia dudosa acompañándola de un “no sé si será cierto, pero…”.

La desinformación no es, por tanto, solo un fallo de conocimiento. También está relacionada con la emoción, la identidad, la confianza, el conflicto entre grupos y los incentivos sociales.

Esto obliga a moderar las expectativas. Enseñar a reconocer una técnica manipuladora puede mejorar el juicio, pero no elimina necesariamente las recompensas emocionales o sociales asociadas a compartir el contenido.

La inoculación parece producir efectos que se mantienen durante cierto tiempo, pero no existe una protección indefinida. La memoria se debilita, aparecen nuevas variantes de la manipulación y el contexto cambia.

Algunos estudios longitudinales indican que los recordatorios o *boosters* pueden recuperar parte del efecto. Esto sugiere que el prebunking debería entenderse como una forma de alfabetización continuada, no como una intervención única.

Aprender una vez qué es un falso dilema no garantiza reconocerlo meses después en una situación emocionalmente intensa. Como ocurre con otras habilidades, la detección de manipulaciones mejora cuando se practica con ejemplos variados y se refuerza periódicamente.

Otra limitación importante afecta a la transferencia de los resultados.

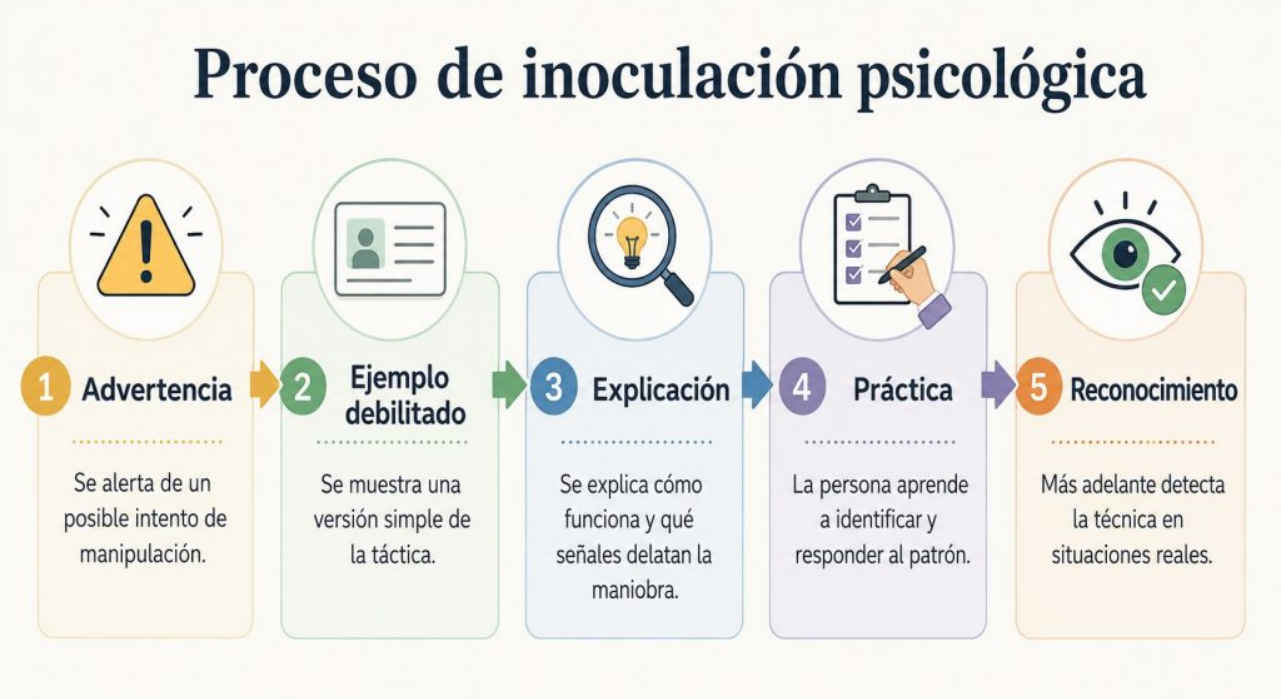
En muchos experimentos, los participantes observan una serie de titulares y deben valorarlos como verdaderos, falsos, fiables o poco fiables. Estas pruebas permiten comparar grupos y controlar variables, pero simplifican considerablemente la vida real.

En una red social intervienen simultáneamente la velocidad de navegación, el cansancio, las notificaciones, la confianza en quien comparte el mensaje, el número de reacciones, la repetición, los comentarios del grupo, el diseño de la interfaz, la recomendación algorítmica y la recompensa inmediata de participar. En esas condiciones, detenerse a analizar una técnica requiere algo más que conocerla.

Los estudios realizados dentro de plataformas reales ofrecen resultados prometedores, pero sigue siendo más fácil demostrar una mejora inmediata en el reconocimiento que un cambio duradero en el comportamiento cotidiano.

La evidencia también indica que las intervenciones no funcionan exactamente igual en todas las culturas, idiomas y contextos. Algunos estudios internacionales han encontrado efectos consistentes, mientras que otros han obtenido resultados más débiles al trasladar las mismas técnicas a poblaciones diferentes.

La conclusión razonable no es que el prebunking funcione siempre ni que fracase cuando [illegible]a, pero



La anatomía de una manipulación

Cuando pensamos en el engaño, solemos concentrarnos en el contenido, en qué dato es falso, qué fotografía se ha modificado o qué declaración se ha inventado.

Sin embargo, muchas manipulaciones eficaces contienen información verdadera. El engaño puede aparecer en la selección, el contexto, el orden, la comparación o la emoción utilizada para presentarla.

Una cifra puede ser correcta y conducir a una conclusión equivocada. Una fotografía puede ser auténtica, pero pertenecer a otro país. Un experto puede poseer una titulación real y hablar fuera de su especialidad. Dos opciones pueden existir, pero no ser las únicas disponibles.

El prebunking basado en técnicas intenta enseñarnos a reconocer estas estructuras.

La emoción no es enemiga de la razón. Nos ayuda a identificar amenazas, priorizar problemas y actuar cuando no disponemos de toda la información. El problema aparece cuando alguien intenta activar una emoción para reducir deliberadamente nuestra capacidad de examinar el mensaje.

Algunas señales habituales son lenguaje extremadamente alarmista, imágenes diseñadas para provocar miedo o asco, indignación moral intensa, uso constante de mayúsculas y exclamaciones, afirmaciones de que “nadie se atreve a decirlo” o amenazas que no se acompañan de datos verificables.

Un mensaje emocional no es necesariamente falso. Una catástrofe real puede provocar miedo y una injusticia auténtica puede generar indignación. La pregunta útil no es si el contenido nos emociona, sino si la emoción que siento me ayuda a comprender la información o está sustituyendo a la evidencia.

La urgencia reduce el tiempo disponible para buscar información, comparar alternativas o consultar a otra persona. Es frecuente en fraudes digitales, ventas agresivas, conflictos laborales, mensajes políticos y cadenas virales.

Expresiones como “ahora o nunca”, “antes de que lo borren” o “última oportunidad” pueden utilizarse para impedir que recuperemos el control sobre el ritmo de la decisión.

La urgencia puede ser legítima. Un incendio, una emergencia médica o un ataque informático requieren actuar con rapidez. La diferencia está en si el plazo puede

justificarse mediante circunstancias verificables o se introduce únicamente para presionar.

Una regla preventiva sencilla sería la siguiente: cuanto mayor sea la presión para decidir, mayor debería ser la necesidad de comprobar por qué existe esa urgencia.

El falso dilema reduce una situación compleja a dos opciones excluyentes tipo “O estás con nosotros o estás contra nosotros”, “O apoyas esta medida o no te preocupa el problema”, “O utilizamos esta tecnología o nos quedamos atrás”.

A veces existen realmente dos alternativas. Pero en muchos problemas sociales, científicos, empresariales o personales hay opciones intermedias, condiciones, combinaciones y grados.

La respuesta inoculadora consiste en preguntar:

- ¿Son realmente las únicas posibilidades?
- ¿Qué alternativas se han omitido?
- ¿Puedo aceptar una parte y rechazar otra?
- ¿Qué condiciones cambiarían mi decisión?

La misma reducción aparece cuando un problema complejo se atribuye a una única persona o grupo. El chivo expiatorio ofrece una causa clara y emocionalmente satisfactoria, pero suele eliminar el contexto.

Que una persona o institución tenga responsabilidad no significa que explique por sí sola todo el fenómeno.

Una de las formas más eficaces de persuasión consiste en presentar a alguien como una autoridad. Basta con añadir una bata blanca, un título académico o una breve referencia a su experiencia para que el mensaje gane credibilidad. Pero la palabra «experto» no debería cerrar la conversación, sino abrir varias preguntas como en qué disciplina se ha formado esa persona, si está hablando realmente dentro de su especialidad, si aporta datos o solo una opinión, si su posición encaja con el conjunto de la evidencia disponible, si existen conflictos de interés y si otros especialistas independientes han llegado a conclusiones parecidas.

Una titulación no convierte todas las afirmaciones de una persona en verdaderas. Un médico puede hablar sin conocimientos suficientes sobre inteligencia artificial, un ingeniero puede opinar sin rigor sobre nutrición, un psicólogo puede interpretar incorrectamente una cuestión epidemiológica. Esto tampoco significa que el consenso sea infalible ni que una minoría esté siempre equivocada. Significa que una opinión aislada

debe valorarse dentro del conjunto de evidencias disponibles.

Por otro lado, los números producen una impresión de objetividad. Precisamente por eso pueden ser herramientas eficaces de manipulación.

Consideremos la frase “Los casos han aumentado un 200 %”. Podría significar que han pasado de uno a tres. Sin conocer el número inicial, el periodo analizado, el tamaño de la población y el método de recogida, el porcentaje tiene poco significado.

Antes de aceptar una cifra conviene preguntar:

- ¿Cuál es el valor absoluto?
- ¿Respecto a qué se compara?
- ¿Durante qué periodo?
- ¿Cuál es el denominador?
- ¿Se habla de riesgo relativo o absoluto?
- ¿La muestra es representativa?
- ¿Se han seleccionado únicamente los años favorables?

Algo parecido ocurre con los estudios científicos. En casi cualquier tema complejo es posible encontrar una investigación, un caso o una experiencia personal que respalde una conclusión determinada. La selección interesada aparece cuando se elige únicamente aquello que confirma una posición y se omite el resto. El antídoto no consiste en buscar otro estudio aislado que diga lo contrario, sino en comprender qué muestra el conjunto de la evidencia.

En la era de los vídeos breves y los fragmentos virales, un contenido no necesita ser falso para engañar. Una declaración auténtica puede recortarse hasta cambiar su significado. Una imagen real puede atribuirse a otro acontecimiento. Una gráfica correcta puede comenzar en un punto elegido para exagerar una diferencia.

La verificación debería incluir:

- localizar la fuente original;
- comprobar la fecha;
- observar qué ocurrió antes y después;
- identificar el lugar;
- buscar la versión completa;
- confirmar cómo se recogieron los datos.

Otra forma de evitar el contenido consiste en atacar a quien lo presenta. La reputación, la competencia y los conflictos de interés de una fuente importan, pero demostrar que alguien es antipático, partidista o interesado no demuestra automáticamente que su argumento sea falso.

La defensa consiste en separar dos preguntas:

1. ¿Es fiable la fuente?
2. ¿Está respaldada la afirmación?

Una fuente poco fiable puede decir algo cierto. Una fuente prestigiosa puede equivocarse.

Algunas teorías se construyen de manera que cualquier posible refutación las confirme. Si no aparecen pruebas, es porque han sido ocultadas. Si aparecen pruebas contrarias, han sido fabricadas. Si un experto discrepa, está comprado. Si una institución investiga, forma parte de la conspiración. Esta estructura convierte la afirmación en prácticamente inmune a la evidencia.

Una pregunta fundamental es ¿Qué prueba podría demostrar que esta afirmación es incorrecta?

Cuando la respuesta es “ninguna”, probablemente no estamos ante una hipótesis verificable, sino ante un sistema cerrado de interpretación.

Cómo aplicar el prebunking en la vida cotidiana

Conocer las técnicas no garantiza que las reconozcamos en el momento adecuado. Necesitamos un procedimiento suficientemente sencillo para utilizarlo cuando la atención es limitada.

Como propuesta práctica propia, yo utilizo un método que puedo convertir para ti, ya que te has tomado la molestia de llegar hasta aquí, en una regla nemotécnica que podríamos llamar el método PAUSA.

No es una escala científica validada, sino un marco de trabajo construido a partir de los mecanismos estudiados en la literatura sobre inoculación, pensamiento crítico y desinformación.

- **P: Para**. Antes de compartir, comprar, contestar o decidir, interrumpe temporalmente la acción. No necesitas realizar una investigación completa. A veces

basta con introducir unos segundos de demora. La pausa resulta especialmente importante cuando el mensaje provoca una emoción intensa, pide actuar inmediatamente, confirma exactamente lo que ya pensabas, ataca a un grupo con el que estás enfrentado, promete un beneficio extraordinario o amenaza con una pérdida inminente. El objetivo no es eliminar la emoción, sino impedir que decida por sí sola.

- **A: Atiende** a la emoción. Pregúntate "¿qué intenta hacerme sentir este mensaje?". Puede intentar generar miedo, rabia, culpa, superioridad, esperanza desproporcionada, exclusividad o deseo de castigar. Después añade una segunda pregunta que te plantee "¿qué ocurriría con mi interpretación si esta emoción disminuyera?". Esta separación ayuda a distinguir entre lo que el mensaje dice y lo que intenta provocar.
- **U: Ubica la técnica.** Intenta reconocer la estrategia utilizada. Podría ser urgencia, falso dilema, enemigo único, experto impostor, cifra sin contexto, evidencia seleccionada, descontextualización, ataque personal o conspiración cerrada. No es imprescindible decidir inmediatamente si el contenido es verdadero o falso, basta con reconocer que está utilizando una técnica capaz de distorsionar el juicio.
- **S: Solicita contexto**. Busca la información necesaria para interpretar la afirmación (fuente original, autor, fecha, datos completos, metodología, comparación adecuada, versión íntegra, opinión de fuentes independientes, explicación alternativa). En muchos casos, una búsqueda lateral resulta más eficaz que permanecer en la propia página. En lugar de leer únicamente lo que una web dice sobre sí misma, podemos consultar quién la publica, qué reputación tiene y cómo describen la información otras fuentes.
- **A: Actúa proporcionalmente**. No todas las afirmaciones merecen el mismo esfuerzo de comprobación. Si alguien comparte una curiosidad sin consecuencias, podemos mantener una duda razonable sin investigar durante una hora. Cuando la decisión afecta a la salud, el dinero, la seguridad, la reputación o los derechos de una persona, el nivel de verificación debe aumentar. Podemos distinguir tres niveles: (1) Riesgo bajo, en cuyo caso conviene no compartir o expresar cautela, (2) Riesgo moderado requiere comprobar la fuente y contrastar con una referencia independiente y (3) Riesgo alto, exige detener la decisión y consultar fuentes especializadas, profesionales competentes o documentación oficial. Pensar críticamente también implica administrar bien nuestra atención. No podemos verificarlo todo, pero sí dedicar más recursos a aquello que puede producir más daño.

Una confusión bastante común consiste en asumir que cuanto más emocional es un mensaje, menos fiable resulta, pero esa relación no es tan sencilla. Una información puede

estar bien documentada y, al mismo tiempo, provocar miedo, tristeza o indignación. Del mismo modo, un contenido falso puede presentarse con apariencia académica, lenguaje técnico y gráficos cuidadosamente diseñados. Una historia personal puede ayudar a comprender un problema real, mientras que una tabla llena de cifras puede esconder una comparación engañosa.

Para evitar esta simplificación conviene separar dos dimensiones distintas: la calidad de la información y el grado de activación emocional que produce. Cuando la calidad es alta y la activación es baja, solemos encontrarnos ante información analítica y contrastable. Si la calidad sigue siendo alta, pero la carga emocional aumenta, estamos ante contenidos relevantes que exigen regular la emoción para poder interpretarlos con claridad. En cambio, una información de baja calidad y poca activación suele convertirse simplemente en ruido. La combinación más peligrosa aparece cuando hay mucha activación emocional y muy poca información verificable, porque es ahí donde la manipulación puede resultar especialmente eficaz.

Esto no significa que debamos desconfiar de todo lo que nos conmueve. Un informe sobre una guerra, una catástrofe o una enfermedad puede estar sólidamente documentado y resultar profundamente perturbador. En esos casos, el problema no está en la emoción, sino en permitir que esa emoción sustituya al análisis. La clave no es rechazar el contenido por cómo nos hace sentir, sino comprobar si lo que sentimos está acompañado por información suficiente, contexto y evidencia.

Del WhatsApp a la empresa y la inteligencia artificial

El prebunking suele estudiarse en relación con la desinformación, pero su lógica puede aplicarse a muchos ámbitos.

En ciberseguridad, una organización puede enseñar previamente a sus empleados qué tácticas utilizará probablemente un atacante, como falsa urgencia, suplantación de un directivo, petición de confidencialidad, cambio inesperado de una cuenta bancaria o solicitud de códigos de acceso.

En una compra importante, podemos anticipar la presión temporal, la escasez artificial, la ocultación de costes o los testimonios seleccionados.

En salud, conviene advertir de patrones como las curas universales, los testimonios presentados como pruebas, los supuestos expertos fuera de su especialidad o los productos naturales descritos como necesariamente seguros.

En una empresa, el prebunking puede utilizarse antes de una decisión relevante, por ejemplo “durante esta reunión podríamos sobrevalorar los datos que confirman el proyecto, minimizar sus costes o interpretar cualquier crítica como falta de compromiso”.

Esto no elimina los sesgos, pero los hace visibles antes de que condicionen la decisión.

La inteligencia artificial generativa amplía todavía más el problema. Hoy es posible producir textos, imágenes, voces y vídeos plausibles con rapidez y a bajo coste. Pero la pregunta no debería ser únicamente si un contenido ha sido creado por una inteligencia artificial.

Un contenido sintético puede ser verdadero, falso, satírico o irrelevante. Del mismo modo, una persona puede crear manualmente una manipulación. La cuestión importante es qué pruebas respaldan lo que afirma y qué intenta conseguir de mí.

El prebunking frente a contenidos generados por IA debería enseñar a reconocer fuentes inexistentes, citas inventadas, autoridad fabricada, testimonios artificiales, imágenes descontextualizadas y simulaciones de consenso.

El lado oscuro del prebunking

Una de las paradojas más importantes de la inoculación psicológica es que la misma técnica puede utilizarse de manera honesta o engañosa.

Imaginemos que una empresa afirma “nuestros competidores intentarán convencerte de que este producto no funciona porque temen nuestra innovación”. La frase prepara al consumidor para interpretar cualquier crítica futura como prueba del interés de los competidores. No aporta evidencia sobre el producto, sino que intenta neutralizar preventivamente la objeción.

Algo parecido ocurre cuando un líder dice “en los próximos días escucharéis acusaciones falsas de personas que quieren destruirnos”. Puede que algunas acusaciones sean falsas, pero la advertencia también puede utilizarse para debilitar la capacidad del grupo de examinar críticas legítimas.

La inoculación no es éticamente buena por definición. Depende de qué se intenta proteger y de cómo se presentan las evidencias.

Una intervención responsable no debería utilizar el prebunking para conducir al público hacia una conclusión previamente decidida, sino para enseñar a reconocer las técnicas de

manipulación. Su función tendría que ser ampliar la autonomía del receptor, no sustituir una influencia por otra.

Eso implica ofrecer herramientas que permitan revisar también las propias creencias, diferenciar la calidad de la evidencia del prestigio de quien habla y reconocer con claridad qué sabemos, qué suponemos y qué continúa siendo incierto. También exige evitar un error frecuente: interpretar cualquier discrepancia como una forma de manipulación. La crítica puede ser interesada, pero también puede ser legítima, necesaria o estar mejor fundamentada que la posición que intenta cuestionar.

Por eso, los criterios que utilizamos para evaluar los mensajes ajenos deberían poder aplicarse igualmente a quien diseña la intervención. Una buena regla ética sería la siguiente:

Una herramienta de pensamiento crítico debe permitir examinar tanto las afirmaciones ajenas como las propias.

El peligro de no creer en nada

La exposición constante a fraudes, bulos, montajes y mensajes interesados puede llevarnos a una respuesta que, en apariencia, parece prudente, es decir, desconfiar de todo. Si cualquiera puede manipular una imagen, inventar una fuente o presentar una opinión como evidencia, quizá lo más seguro sea no creer en nada.

Sin embargo, el escepticismo indiscriminado no mejora nuestro conocimiento. Si concedemos la misma credibilidad a una revisión científica, un vídeo anónimo, una declaración institucional y una cadena de WhatsApp, no estamos pensando críticamente. Estamos renunciando a distinguir entre fuentes que ofrecen niveles de evidencia muy diferentes.

El objetivo no debería ser la desconfianza permanente, sino la calibración de la confianza. Esto significa aumentar nuestra confianza cuando distintas pruebas independientes convergen, reducirla cuando no conocemos la fuente o faltan datos esenciales y mantener cierta incertidumbre cuando la evidencia todavía es incompleta. También significa estar dispuestos a revisar una posición cuando aparecen nuevos resultados y no confundir la seguridad con la que alguien habla con la solidez de lo que afirma.

La evidencia disponible sugiere que el prebunking puede mejorar el discernimiento sin convertir necesariamente a las personas en desconfiadas sistemáticas. Pero eso no

garantiza que cualquier campaña de inoculación produzca ese resultado. Una intervención mal diseñada puede fomentar cinismo, reforzar la polarización o alimentar el rechazo indiscriminado hacia instituciones, medios de comunicación y expertos.

Conclusión

El prebunking puede resultar más eficaz para enseñar la anatomía de una manipulación antes de que nos alcance que para intentar corregir cada engaño después de que haya influido sobre nosotros.

La evidencia respalda parcialmente esta idea. Las intervenciones de inoculación psicológica pueden mejorar el discernimiento y ayudarnos a reconocer técnicas como la urgencia artificial, el falso dilema, la apelación emocional, la autoridad impostada o la descontextualización.

Pero no existe una vacuna mental definitiva.

Podemos conocer una técnica y no detectarla cuando confirma nuestras creencias. Podemos reconocer que un contenido es dudoso y compartirlo porque expresa nuestra identidad. Podemos defendernos de una forma de manipulación y caer en otra más sofisticada.

Además, el prebunking no resolverá por sí solo los incentivos económicos de la desinformación, la opacidad algorítmica, las campañas coordinadas o la pérdida de confianza institucional. No deberíamos trasladar toda la responsabilidad a una persona que compite por su atención dentro de sistemas diseñados para capturarla.

Su valor es más modesto, pero no menor.

El prebunking puede enseñarnos a reconocer cuándo alguien intenta reducir nuestro espacio para pensar. Puede ayudarnos a recuperar el tiempo de la decisión, separar la emoción de la evidencia y pedir el contexto que falta.

No nos protege únicamente contra afirmaciones falsas, sino contra los intentos de alterar las condiciones bajo las que decidimos.

Evitar que nos engañen no consiste en no creer en nada. Consiste en saber por qué confiamos, qué pruebas necesitaríamos para cambiar de opinión y cuándo una emoción, una urgencia o una autoridad aparente están intentando decidir por nosotros.

No seremos inmunes, pero quizá aprendamos a detenernos antes de reaccionar. Y, en un entorno que compite constantemente por nuestra atención, esa pausa puede ser una de las formas más valiosas de libertad.

Referencias

- Banas, J. A., & Rains, S. A. (2010). A meta-analysis of research on inoculation theory. *Communication Monographs, 77*(3), 281–311. https://doi.org/10.1080/03637751003758193
- Compton, J., van der Linden, S., Cook, J., & Basol, M. (2021). Inoculation theory in the post-truth era: Extant findings and new frontiers for contested science, misinformation, and conspiracy theories. *Social and Personality Psychology Compass, 15*(6), e12602. https://doi.org/10.1111/spc3.12602
- Cook, J., Lewandowsky, S., & Ecker, U. K. H. (2017). Neutralizing misinformation through inoculation: Exposing misleading argumentation techniques reduces their influence. *PLOS ONE, 12*(5), e0175799. https://doi.org/10.1371/journal.pone.0175799
- Ecker, U. K. H., Lewandowsky, S., Cook, J., Schmid, P., Fazio, L. K., Brashier, N. M., Kendeou, P., Vraga, E. K., & Amazeen, M. A. (2022). The psychological drivers of misinformation belief and its resistance to correction. *Nature Reviews Psychology, 1*, 13–29. https://doi.org/10.1038/s44159-021-00006-y
- Lewandowsky, S., & van der Linden, S. (2021). Countering misinformation and fake news through inoculation and prebunking. *European Review of Social Psychology, 32* (2), 348–384. https://doi.org/10.1080/10463283.2021.1876983
- Lu, C., Hu, B., Li, Q., Bi, C., & Ju, X. D. (2023). Psychological inoculation for credibility assessment, sharing intention, and discernment of misinformation: Systematic review and meta-analysis. *Journal of Medical Internet Research, 25*, e49255. https://doi.org/10.2196/49255
- Maertens, R., Roozenbeek, J., Basol, M., & van der Linden, S. (2021). Long-term effectiveness of inoculation against misinformation: Three longitudinal experiments. *Journal of Experimental Psychology: Applied, 27*(1), 1–16. https://doi.org/10.1037/xap0000315
- Roozenbeek, J., van der Linden, S., Goldberg, B., Rathje, S., & Lewandowsky, S. (2022). Psychological inoculation improves resilience against misinformation on social media. *Science Advances, 8*(34), eabo6254. https://doi.org/10.1126/sciadv.abo6254
- Roozenbeek, J., & van der Linden, S. (2019). Fake news game confers psychological resistance against online misinformation. *Palgrave Communications, 5*, 65. https://doi.org/10.1057/s41599-019-0279-9

Cómo evitar engaños

Qué es el prebunking y cómo aplicar el método PAUSA

1 ¿Qué es el prebunking?

El **prebunking**, o inoculación psicológica, consiste en prepararnos antes de encontrarnos con un mensaje engañoso. En lugar de corregir una mentira después, nos enseña a reconocer las tácticas de manipulación antes de caer en ellas.

2 Engaños frecuentes

Urgencia artificial
Te presionan para actuar sin pensar.

Miedo
Intentan activar alarma o angustia.

Autoridad impostada
Usan supuestos expertos para convencer.

Cifras sin contexto
Dan porcentajes llamativos sin explicar su base.

Falso dilema
Presentan solo dos opciones cuando hay más.

Descontextualización
Usan datos, frases o imágenes fuera de contexto.

3 Método PAUSA

4 Idea clave

> **Evitar un engaño no consiste en desconfiar de todo, sino en aprender a pensar antes de reaccionar.**

del blog **www.sergiocolado.com**

ventions to
8/s41591-