FUNDAMEDIOS

VERSIÓN COMPLETA · 2026

Manual de Adopción y Uso de

Inteligencia Artificial



La IA apoya,
la responsabilidad es humana.

ECUADOR Chequea PERIODISMO CON RIGOR | Liberties 360°

Ilustraciones generadas con IA

CONTENIDO

Índice

Presentación, objetivo, ámbito de aplicación y uso del Manual		**3**
I	Marco general para el uso de inteligencia artificial	**6**
II	Investigación y levantamiento de información mediante inteligencia artificial	**24**
III	Gestión, almacenamiento y protección de la información	**49**
IV	Uso de inteligencia artificial para el análisis y producción de contenidos	**88**
V	Gobernanza institucional del sistema de inteligencia artificial	**138**
VI	Ciberseguridad, gestión de incidentes y continuidad	**148**
VII	Cumplimiento, responsabilidades y disposiciones finales	**161**

Anexos

1	Declaración de uso responsable de inteligencia artificial	**174**
2	Matriz de clasificación de información para uso con inteligencia artificial	**181**
3	Matriz institucional de riesgos para el uso de inteligencia artificial	**191**
4	Lista de comprobación para investigación y levantamiento de información mediante inteligencia artificial	**202**
5	Lista de comprobación para producción y publicación asistida por inteligencia artificial	**216**
6	Formato de registro de incidentes relacionados con inteligencia artificial	**232**
7	Catálogo de herramientas autorizadas para uso institucional	**245**
8	Biblioteca inicial de prompts institucionales	**255**
9	Ciclo institucional para el uso de inteligencia artificial	**273**

FUNDAMEDIOS

VERSIÓN COMPLETA · 2026

Manual de Adopción y Uso de Inteligencia Artificial

Manual para el uso responsable, ético y seguro de inteligencia artificial

Fundamedios, Ecuador Chequea, Liberties 360° y demás proyectos, programas y actividades desarrollados por Fundamedios

Presentación

La inteligencia artificial se ha incorporado progresivamente como una herramienta de apoyo para la búsqueda, procesamiento, análisis y producción de información. Su utilización puede contribuir significativamente a las actividades periodísticas, investigativas, de verificación, monitoreo, comunicación y gestión desarrolladas por Fundamedios y sus distintos proyectos.

Estas herramientas permiten procesar grandes volúmenes de información, localizar y organizar fuentes, analizar documentos, identificar patrones, sistematizar datos, generar borradores y asistir en diferentes etapas de la producción de contenidos. Sin embargo, su utilización también genera riesgos relacionados con la exactitud de la información, la generación de información inexistente o incorrecta, los sesgos, la protección de datos personales, la confidencialidad, la propiedad intelectual, la seguridad de la información y la determinación de responsabilidades.

Fundamedios adopta este Manual con el propósito de establecer un marco institucional común para incorporar herramientas de inteligencia artificial a sus actividades sin reducir sus estándares periodísticos, investigativos, editoriales, éticos y de seguridad.

El Manual parte de una premisa fundamental: **la inteligencia artificial constituye una herramienta de asistencia al trabajo humano y no sustituye el criterio profesional ni la responsabilidad de las personas que investigan, verifican, analizan, redactan, revisan o autorizan la publicación de información.**

Su utilización deberá, por tanto, integrarse a procedimientos que permitan mantener control humano sobre las distintas etapas del tratamiento de información y de la producción de contenidos.

Objetivo

El presente Manual tiene por objeto establecer los criterios, procedimientos y controles institucionales aplicables al uso de herramientas y sistemas de inteligencia artificial en las actividades desarrolladas por Fundamedios.

En particular, busca establecer un procedimiento común para el ciclo completo de utilización de inteligencia artificial:

> **planificación → levantamiento de información → registro → clasificación → protección → verificación → análisis → producción → revisión humana → aprobación → publicación → conservación.**

Este procedimiento busca garantizar que la utilización de inteligencia artificial se realice bajo condiciones de:

- exactitud y verificabilidad;
- transparencia y trazabilidad;

- supervisión humana;
- protección de datos personales;
- confidencialidad;
- seguridad de la información;
- responsabilidad;
- respeto de los estándares periodísticos y editoriales de la organización; y
- cumplimiento de la normativa aplicable.

El Manual busca, además, que cuando se produzca un error, incidente, reclamación o controversia relacionada con un producto elaborado con asistencia de inteligencia artificial, sea posible reconstruir el procedimiento seguido, identificar las fuentes utilizadas, determinar los controles aplicados y establecer las responsabilidades correspondientes.

Ámbito de aplicación

El presente Manual será aplicable al uso de inteligencia artificial realizado en el marco de las actividades de **Fundamedios, Ecuador Chequea, Liberties 360°**, así como de cualquier otro proyecto, programa, plataforma, producto, iniciativa o denominación administrada o desarrollada por Fundamedios.

Será de aplicación para todas las personas que utilicen herramientas de inteligencia artificial en actividades relacionadas con la organización, incluyendo personal directivo, periodistas, editores, investigadores, verificadores, analistas, personal técnico y administrativo, consultores, colaboradores, pasantes y, cuando corresponda, proveedores o terceros que procesen información por cuenta de Fundamedios.

El Manual se aplicará tanto a las herramientas de inteligencia artificial proporcionadas o contratadas institucionalmente como a aquellas utilizadas por una persona para desarrollar actividades o productos destinados a Fundamedios o cualquiera de sus proyectos.

Comprende el uso de inteligencia artificial para actividades de investigación, búsqueda y levantamiento de información, monitoreo, verificación, análisis documental o de datos, clasificación y sistematización de información, transcripción, traducción, elaboración de resúmenes, generación y edición de textos, producción de contenidos audiovisuales, elaboración de alertas, notas, informes y demás productos institucionales o destinados al público.

Cómo utilizar este manual

Este documento ha sido concebido como un **manual operativo**. Sus capítulos siguen, en la medida de lo posible, el recorrido que realiza la información dentro de Fundamedios cuando intervienen herramientas de inteligencia artificial.

El Manual distingue dos grandes momentos del proceso:

1. Obtención y tratamiento de información. Comprende la búsqueda, recolección, registro, clasificación, almacenamiento, protección, análisis y verificación de información mediante herramientas de inteligencia artificial.

2. Producción y publicación. Comprende la utilización de información previamente obtenida y validada para elaborar, con asistencia de inteligencia artificial, alertas, notas periodísticas, verificaciones, informes, contenidos digitales y demás productos de Fundamedios.

Entre ambos momentos deberá existir siempre el nivel de **verificación y supervisión humana** que corresponda a la naturaleza y riesgo de la información.

Los capítulos siguientes desarrollan las reglas y procedimientos aplicables a cada etapa y establecen los controles necesarios para que la utilización de inteligencia artificial pueda ser documentada, supervisada y, cuando resulte necesario, auditada.

CAPÍTULO

I

Marco general para el uso de inteligencia artificial



1. Principios generales para el uso de inteligencia artificial

El uso de herramientas de inteligencia artificial en Fundamedios deberá estar orientado a fortalecer las capacidades de investigación, análisis, verificación y producción de contenidos de la organización, sin sustituir los estándares profesionales, periodísticos, editoriales, éticos y de seguridad aplicables a cada actividad.

Toda persona que utilice inteligencia artificial en actividades de Fundamedios deberá observar los siguientes principios durante el levantamiento de información, su procesamiento y análisis, la elaboración de contenidos y su eventual publicación.

1.1. Supervisión humana

La inteligencia artificial será utilizada como una **herramienta de asistencia al trabajo humano**.

Toda información, análisis o contenido generado mediante IA que pueda incidir sustancialmente en un producto de Fundamedios deberá estar sujeto a supervisión humana antes de ser utilizado o publicado.

La persona responsable deberá estar en capacidad de revisar, cuestionar, corregir o descartar el resultado producido por la herramienta. La revisión no deberá limitarse a comprobar que el texto se encuentre correctamente redactado, sino que deberá considerar la exactitud de la información, la calidad y correspondencia de las fuentes, el contexto, las conclusiones y los posibles riesgos derivados de su utilización.

La decisión final sobre la utilización y publicación de información será siempre humana.

1.2. Exactitud y verificación

Los resultados proporcionados por una herramienta de inteligencia artificial **no deberán presumirse correctos por el solo hecho de haber sido generados por el sistema**.

La información factual relevante obtenida mediante IA deberá ser verificada antes de ser incorporada a un producto destinado a publicación.

Especial atención deberá prestarse a nombres, fechas, cifras, citas, antecedentes, estadísticas, referencias documentales, decisiones judiciales, disposiciones normativas y demás información cuya alteración pueda modificar el sentido de una publicación.

Una respuesta generada por inteligencia artificial no constituye, por sí misma, una fuente. Cuando la IA permita localizar determinada información, deberá identificarse y, cuando sea razonablemente posible, consultarse la fuente que sustenta dicha información.

No deberá publicarse como hecho comprobado una afirmación cuya única sustentación sea el resultado producido por una herramienta de inteligencia artificial.

1.3. Calidad de la información

La calidad del resultado producido mediante inteligencia artificial depende, entre otros factores, de la calidad de la información utilizada como punto de partida.

En los procesos de investigación y análisis deberán privilegiarse datos y fuentes que sean pertinentes, suficientemente completas, actualizadas y adecuadas para la finalidad perseguida.

La utilización de grandes cantidades de información no sustituye la evaluación de su calidad.

Cuando existan indicios de que los datos utilizados pueden encontrarse incompletos, desactualizados, descontextualizados o afectados por sesgos, esta circunstancia deberá ser considerada antes de utilizar los resultados obtenidos mediante IA.

1.4. Trazabilidad

Cuando la inteligencia artificial tenga una intervención sustancial en la obtención, análisis o producción de información, deberá existir un registro suficiente para reconstruir las principales etapas del proceso.

La trazabilidad deberá permitir determinar, según corresponda:

> **qué se buscó → mediante qué herramienta → qué información o fuentes se obtuvieron → cuáles fueron utilizadas → cómo fueron verificadas → cómo fueron procesadas → quién realizó la revisión → qué información llegó al producto final.**

El nivel de registro será proporcional a la naturaleza y riesgo de la actividad. No será necesario documentar de igual manera una corrección ortográfica que una investigación periodística, una verificación de hechos o una alerta relacionada con posibles vulneraciones de derechos.

La trazabilidad tendrá como finalidad facilitar la revisión editorial, la corrección de errores, la atención de reclamaciones, la investigación de incidentes y, cuando corresponda, la determinación de responsabilidades.

1.5. Transparencia

La utilización de inteligencia artificial deberá realizarse de manera transparente tanto dentro de la organización como, cuando corresponda por la naturaleza de su intervención, frente al público.

Las personas que participen en un proceso deberán poder identificar cuándo la IA tuvo una intervención sustancial en la investigación, análisis o producción de un contenido.

Cuando el grado o naturaleza de la intervención de inteligencia artificial resulte relevante para comprender cómo fue producido determinado contenido, se aplicarán los mecanismos de transparencia o disclosure establecidos en este Manual.

La transparencia no implicará revelar información confidencial, fuentes protegidas, metodologías cuya divulgación pueda comprometer investigaciones ni información cuya difusión pueda generar riesgos de seguridad.

1.6. Protección de datos personales y minimización

La utilización de inteligencia artificial deberá respetar la normativa aplicable en materia de protección de datos personales y los procedimientos internos de Fundamedios.

Solo deberán introducirse, recopilarse o procesarse mediante herramientas de IA aquellos datos personales que resulten necesarios y pertinentes para la finalidad legítima que se persigue.

Antes de proporcionar información personal a una herramienta de inteligencia artificial deberá evaluarse si la misma actividad puede realizarse utilizando menos información o suprimiendo elementos que permitan identificar innecesariamente a una persona.

Cuando sea posible y adecuado para la finalidad perseguida, deberán emplearse mecanismos de anonimización, seudonimización, reducción o exclusión de datos personales.

La circunstancia de que determinada información sea públicamente accesible no elimina, por sí sola, la necesidad de evaluar las condiciones bajo las cuales puede ser recopilada, combinada, almacenada o procesada mediante inteligencia artificial.

1.7. Confidencialidad

La información confidencial, reservada o protegida no deberá ser introducida en herramientas de inteligencia artificial que no hayan sido autorizadas para procesar ese nivel de información.

Deberá prestarse especial protección a la identidad y datos de fuentes confidenciales, denunciantes, víctimas y personas en situación de riesgo; documentos internos reservados; información no publicada; credenciales y mecanismos de acceso; información protegida por secreto profesional; estrategias jurídicas; investigaciones en curso; y cualquier información cuya divulgación pueda perjudicar a una persona, una investigación o a la organización.

El usuario deberá considerar que proporcionar información a una herramienta externa de IA constituye una forma de procesamiento de esa información por un tercero, salvo que las condiciones técnicas y contractuales aplicables garanticen un entorno institucional controlado.

1.8. Seguridad de la información

El uso de inteligencia artificial deberá realizarse mediante herramientas, cuentas, dispositivos y mecanismos compatibles con los estándares de seguridad establecidos por Fundamedios.

La protección deberá abarcar todo el ciclo de información: los datos introducidos en una herramienta, archivos adjuntos, instrucciones o prompts, resultados generados, documentos descargados, registros del proceso y contenidos almacenados posteriormente.

No deberán incorporarse contraseñas, claves de acceso, tokens, credenciales u otros mecanismos de autenticación dentro de prompts o archivos proporcionados a herramientas de inteligencia artificial.

Los incidentes que puedan comprometer información institucional o datos personales deberán ser comunicados inmediatamente mediante el procedimiento establecido en este Manual.

1.9. Finalidad y proporcionalidad

La inteligencia artificial deberá utilizarse para una finalidad profesional determinada, legítima y relacionada con las actividades de Fundamedios.

La herramienta utilizada, la cantidad de información procesada y el nivel de automatización deberán guardar proporción con la finalidad perseguida.

No deberá recopilarse o procesarse información de manera indiscriminada simplemente porque una herramienta permita hacerlo.

Cuando puedan alcanzarse los mismos objetivos mediante un tratamiento menos invasivo, especialmente respecto de datos personales o información sensible, deberá preferirse esa alternativa.

1.10. Responsabilidad

La utilización de una herramienta de inteligencia artificial no elimina la responsabilidad de las personas que intervienen en una investigación, análisis, redacción, revisión o publicación.

Cada persona será responsable de cumplir los controles que le correspondan dentro del proceso y de no utilizar resultados generados por IA como sustituto de las verificaciones exigidas por este Manual.

La existencia de un error generado inicialmente por una herramienta de inteligencia artificial no eximirá automáticamente de responsabilidad cuando el procedimiento aplicable exigía su detección mediante revisión humana.

De igual manera, la existencia de un resultado incorrecto producido por IA no determinará automáticamente responsabilidad individual. Ante un incidente deberá revisarse el proceso seguido, las instrucciones proporcionadas, las fuentes disponibles, los controles realizados y el cumplimiento de las obligaciones asignadas a cada participante.

1.11. Precaución frente a resultados inciertos

Cuando existan dudas razonables sobre la exactitud de un resultado, la procedencia de una fuente, la legitimidad del tratamiento de determinados datos, la seguridad de una herramienta o las consecuencias de utilizar determinada información, el resultado no deberá avanzar dentro del proceso hasta que la duda sea razonablemente resuelta.

La ausencia de certeza deberá conducir a una mayor verificación, no a una menor.

La rapidez, facilidad o aparente calidad del resultado generado por una herramienta de inteligencia artificial no justificará reducir los estándares de investigación, verificación, protección de datos, confidencialidad, seguridad o revisión editorial de Fundamedios.

---

2. Modelo de gestión de riesgos para el uso de inteligencia artificial

Fundamedios aplicará un enfoque basado en riesgos para determinar las condiciones bajo las cuales podrá utilizarse inteligencia artificial en sus distintas actividades.

El nivel de control no dependerá únicamente de la herramienta utilizada, sino principalmente de **qué se pretende hacer con ella, qué información será procesada y qué consecuencias podría producir un error, una filtración o un uso inadecuado**.

Por ello, antes de utilizar inteligencia artificial en una actividad sustantiva deberá considerarse el nivel de riesgo del uso previsto. A mayor riesgo deberán existir mayores controles de acceso, protección de información, trazabilidad, verificación y supervisión humana.

2.1. Factores para determinar el nivel de riesgo

La evaluación deberá considerar conjuntamente, según corresponda:

- la finalidad para la cual se utilizará la IA;
- la naturaleza de la información que será introducida o procesada;
- la existencia de datos personales, datos sensibles o información confidencial;
- la confiabilidad y procedencia de las fuentes utilizadas;
- la posibilidad de que el sistema produzca información falsa, incompleta, descontextualizada o sesgada;
- el grado de intervención de la IA en el análisis o producto final;
- el nivel de autonomía otorgado al sistema;
- la posibilidad y facilidad de realizar una verificación humana;
- las personas que podrían resultar afectadas;
- la posible afectación a la honra, reputación, privacidad, seguridad u otros derechos;
- la sensibilidad periodística, jurídica, política o institucional del asunto;
- la posibilidad de revelar fuentes periodísticas, investigaciones en curso o información no publicada;
- el alcance previsto de la difusión del producto; y
- la gravedad y reversibilidad de las consecuencias que podría producir un error.

La presencia de uno de estos factores no determinará necesariamente por sí sola el nivel de riesgo. Deberá evaluarse el contexto completo de la actividad.

2.2. Riesgo bajo

Se considerarán de **riesgo bajo** aquellos usos de inteligencia artificial que tengan una función principalmente instrumental y cuya utilización incorrecta presente una probabilidad reducida de causar perjuicios relevantes.

Podrán incluir, entre otros:

- corrección ortográfica y gramatical;
- adecuación de formato;
- organización de información no sensible;
- generación de índices;
- conversión de información previamente validada a tablas o estructuras;
- propuestas de títulos;
- lluvia de ideas;
- apoyo para estructurar documentos;
- traducciones preliminares de información no sensible; y
- resumen de documentos previamente conocidos por el usuario cuando el resultado no vaya a utilizarse como fuente factual autónoma.

Estos usos podrán realizarse con controles simplificados.

No obstante, si durante una actividad aparentemente de bajo riesgo se incorporan datos personales sensibles, información confidencial, fuentes protegidas o información cuya alteración pueda producir consecuencias relevantes, el uso deberá reclasificarse.

2.3. Riesgo medio

Se considerarán de **riesgo medio** los usos en los cuales la inteligencia artificial intervenga materialmente en el levantamiento, clasificación, análisis o producción de información que posteriormente pueda formar parte de una actividad institucional o contenido público.

Podrán incluir:

- búsquedas asistidas de información;
- identificación de posibles fuentes;
- análisis inicial de documentos públicos;
- comparación de documentos;
- elaboración de cronologías;
- clasificación de grandes cantidades de información;
- identificación de posibles inconsistencias;
- extracción estructurada de datos;
- generación de resúmenes destinados a apoyar una investigación;
- preparación de borradores de contenidos a partir de información previamente validada; y
- asistencia en la preparación de alertas, notas, informes u otros productos que posteriormente serán sometidos a revisión humana.

En estos casos deberá existir, como mínimo, identificación de las fuentes relevantes, verificación humana de los elementos factuales utilizados y revisión del resultado antes de que pueda incorporarse a un producto final.

2.4. Riesgo alto

Se considerarán de **riesgo alto** aquellos usos en los cuales un error, sesgo, filtración, inferencia incorrecta o utilización inadecuada de inteligencia artificial pueda generar consecuencias significativas para una persona, una investigación, la organización o la credibilidad de sus productos.

Podrán considerarse de riesgo alto, entre otros:

- investigaciones sobre posibles actos ilícitos o irregularidades;
- contenidos que atribuyan hechos o conductas potencialmente perjudiciales a personas identificadas o identificables;
- verificaciones cuyo resultado pueda calificar públicamente una afirmación como falsa, engañosa o incorrecta;
- investigaciones relacionadas con corrupción, crimen organizado, violencia, amenazas o vulneraciones de derechos;
- procesamiento de información relativa a víctimas, denunciantes o personas en situación de vulnerabilidad;
- análisis de grandes conjuntos de datos personales;
- utilización de IA para establecer relaciones, patrones o inferencias respecto de personas;
- análisis de información cuya interpretación requiera conocimientos jurídicos, financieros, científicos o técnicos especializados;
- utilización de IA para analizar imágenes, audio o video cuando de dicho análisis pueda derivarse una afirmación factual relevante;
- generación o modificación mediante IA de imágenes, audio o video destinados al público; y
- cualquier utilización en la cual un resultado incorrecto pueda generar una afectación jurídica, reputacional, económica o de seguridad relevante.

En estos casos deberá existir **supervisión humana reforzada**.

Los hechos centrales deberán verificarse directamente en las fuentes correspondientes. Las conclusiones generadas o sugeridas por IA no podrán utilizarse sin evaluación humana independiente y deberá conservarse la trazabilidad necesaria para reconstruir el proceso.

Cuando la naturaleza del contenido lo requiera, deberá intervenir adicionalmente el responsable editorial, jurídico, de protección de datos o de seguridad correspondiente.

2.5. Riesgo crítico

Se considerará de **riesgo crítico** aquel uso de inteligencia artificial que pueda comprometer gravemente derechos, seguridad, confidencialidad o integridad de las personas o de la organización, o cuya utilización resulte incompatible con los principios establecidos en este Manual.

Entre otros, se considerarán usos críticos:

- introducir en herramientas no autorizadas la identidad de fuentes periodísticas confidenciales;
- proporcionar contraseñas, credenciales, claves de acceso o información que permita comprometer sistemas institucionales;
- utilizar IA para adoptar autónomamente decisiones editoriales sobre acusaciones graves contra personas;
- publicar información generada por IA como verdadera cuando no haya sido posible verificarla;
- fabricar citas, testimonios, documentos, imágenes, audio o video y presentarlos como auténticos;
- utilizar IA para suplantar personas o fuentes;
- emplear sistemas de IA para manipular deliberadamente información destinada al público;

- procesar información altamente sensible o confidencial mediante herramientas que no ofrezcan garantías suficientes para ese tratamiento; y
- realizar cualquier utilización de IA que sea ilícita o manifiestamente incompatible con los estándares periodísticos y éticos de Fundamedios.

Los usos críticos **no podrán realizarse**, salvo que el riesgo que origina la clasificación pueda eliminarse previamente mediante un procedimiento autorizado que transforme la actividad en una categoría admisible.

2.6. Reclasificación del riesgo

La clasificación deberá mantenerse durante todo el proceso.

Una actividad podrá comenzar como de bajo o medio riesgo y convertirse posteriormente en una actividad de riesgo alto debido a la información encontrada, las personas involucradas o el uso que finalmente vaya a darse al resultado.

Por ejemplo, una búsqueda asistida inicialmente destinada a localizar información pública puede adquirir un nivel de riesgo superior cuando revela datos personales sensibles, identifica una fuente confidencial o comienza a utilizarse para sustentar una imputación respecto de una persona.

Cuando cambien las circunstancias relevantes, deberán aplicarse inmediatamente los controles correspondientes al nuevo nivel de riesgo.

2.7. Relación entre riesgo y control humano

La clasificación determinará la intensidad de los controles aplicables:

Nivel	Uso de IA	Verificación humana	Trazabilidad	Aprobación
Bajo	Instrumental o auxiliar	Ordinaria	Simplificada	Flujo habitual
Medio	Participación material	Obligatoria sobre información relevante	Obligatoria respecto de fuentes y resultados utilizados	Revisión del responsable del producto
Alto	Incidencia significativa o posibles afectaciones relevantes	Reforzada e independiente respecto de los elementos centrales	Completa respecto del proceso sustancial	Aprobación editorial y, cuando corresponda, revisión especializada
Crítico	Uso incompatible o riesgo grave no controlado	No habilita el uso	—	Uso prohibido mientras subsista el riesgo

Esta clasificación no sustituye los procedimientos específicos establecidos en los capítulos siguientes. Su función es determinar **cuánta precaución, documentación, verificación y supervisión deberá acompañar cada utilización de inteligencia artificial**.

2.8. Regla de prevalencia del mayor riesgo

Cuando una misma actividad presente características correspondientes a diferentes categorías, se aplicará el nivel de riesgo más alto.

Cuando existan dudas razonables entre dos niveles, se aplicará provisionalmente el nivel superior hasta que una revisión permita determinar que los riesgos se encuentran suficientemente controlados.

La clasificación no deberá utilizarse como un trámite formal destinado únicamente a habilitar el uso de una herramienta. Su finalidad será identificar anticipadamente los riesgos y determinar los controles necesarios para prevenirlos.

3. Roles, responsabilidades y niveles de supervisión

La incorporación de inteligencia artificial a las actividades de Fundamedios no altera la estructura ordinaria de responsabilidad sobre sus investigaciones, verificaciones, alertas, publicaciones y demás productos. La IA se integra como una herramienta dentro de esos procesos y, por tanto, su utilización queda sometida a los niveles de revisión y aprobación que correspondan en cada caso.

Para efectos de este Manual se distinguen las funciones de **usuario u operador, responsable del producto, revisor humano y aprobador editorial**. A ellas podrán sumarse revisiones especializadas cuando la naturaleza o el riesgo de una actividad así lo requieran.

Estas funciones no representan necesariamente cargos o puestos distintos dentro de la organización. Una misma persona podrá desempeñar más de una función cuando el nivel de riesgo lo permita.

3.1. Usuario u operador de inteligencia artificial

Es la persona que interactúa directamente con una herramienta de inteligencia artificial para desarrollar una actividad relacionada con Fundamedios.

Le corresponde:

a) Uso adecuado de la herramienta. Utilizar herramientas autorizadas y respetar las restricciones establecidas para cada una de ellas, especialmente en función de la información que vaya a procesarse.

b) Definición de la tarea. Determinar con claridad la finalidad del uso y proporcionar al sistema instrucciones adecuadas para el trabajo requerido.

c) Protección de la información. Evitar la incorporación de datos personales, información confidencial, documentos reservados o cualquier otro contenido cuyo procesamiento no se encuentre permitido en la herramienta utilizada.

d) Control de fuentes. Identificar las fuentes relevantes obtenidas durante el proceso y distinguir claramente entre información proveniente de una fuente y contenido generado, resumido, interpretado o inferido por la IA.

e) Verificación. No considerar como comprobada una información por el solo hecho de haber sido proporcionada por el sistema y realizar las verificaciones que correspondan antes de utilizarla.

f) Registro. Conservar la información necesaria para asegurar la trazabilidad del proceso cuando así lo exija el nivel de riesgo.

g) Escalamiento. Informar oportunamente los errores, resultados inesperados, dudas o incidentes que no puedan ser resueltos dentro de sus atribuciones.

El operador deberá mantener una actitud crítica frente a los resultados de la herramienta. La coincidencia entre una respuesta generada por IA y la hipótesis o expectativa previa del usuario no constituye una forma de verificación.

3.2. Responsable del producto

Toda investigación, alerta, verificación, nota, informe u otro producto sustantivo elaborado con una intervención relevante de inteligencia artificial deberá tener una persona responsable.

El responsable del producto deberá asegurarse de que la utilización de IA se haya realizado conforme a este Manual y de que el resultado conserve los estándares profesionales y editoriales aplicables.

Su revisión comprenderá principalmente cuatro aspectos:

Primero, las fuentes. Deberá comprobar que los hechos centrales se encuentren respaldados por fuentes identificables y que las fuentes utilizadas sean adecuadas para sostener las afirmaciones realizadas.

Segundo, la verificación. Deberá comprobar que los elementos factuales relevantes hayan sido efectivamente contrastados y que el contenido no reproduzca como hechos inferencias, suposiciones o alucinaciones del sistema.

Tercero, la protección de la información. Deberá verificar que durante el proceso se hayan observado las reglas aplicables sobre datos personales, confidencialidad y seguridad.

Cuarto, la integridad del producto. Deberá comprobar que exista correspondencia entre las fuentes, el análisis realizado y las conclusiones que finalmente se presentan.

Cuando el responsable del producto sea también quien utilizó directamente la herramienta de IA, deberá realizar una revisión diferenciada del resultado generado antes de incorporarlo como trabajo propio. El hecho de haber elaborado personalmente los prompts o dirigido el proceso no sustituye esta revisión.

3.3. Revisor humano

La revisión humana constituye un control sustantivo y no una formalidad previa a la publicación.

Quien revise un producto deberá aproximarse críticamente al trabajo realizado y estar en condiciones de cuestionar, modificar o rechazar tanto los resultados producidos por la IA como las decisiones adoptadas durante su utilización.

La revisión deberá concentrarse, según la naturaleza del producto, en:

a) la correspondencia entre las afirmaciones y sus fuentes;

b) la exactitud de nombres, fechas, cifras, citas y demás datos relevantes;

c) la suficiencia y calidad de la verificación realizada;

d) la existencia de información omitida que pueda alterar sustancialmente el contexto;

e) las inferencias o conclusiones que no se desprendan razonablemente de las fuentes;

f) la presencia de sesgos o formulaciones que puedan distorsionar los hechos;

g) el tratamiento adecuado de datos personales e información confidencial;

h) las posibles afectaciones a derechos de terceros; y

i) el cumplimiento de los estándares editoriales correspondientes al producto.

En los casos de riesgo alto, los hechos y conclusiones centrales deberán recibir una segunda revisión humana siempre que la estructura y disponibilidad de personal lo permitan.

3.4. Aprobación editorial

La decisión de publicar un contenido corresponde exclusivamente a las personas que tengan atribuida esa responsabilidad dentro de Fundamedios, Ecuador Chequea, Liberties 360° o el proyecto respectivo.

Ningún sistema de inteligencia artificial podrá sustituir la aprobación editorial.

Cuando la IA haya tenido una intervención sustancial, quien autorice la publicación deberá contar con información suficiente para comprender cómo fue utilizada. Dependiendo del nivel de riesgo, deberá poder conocer la finalidad

de su utilización, las fuentes principales, las verificaciones efectuadas, la persona responsable de la revisión y cualquier advertencia o limitación que permanezca pendiente.

La aprobación editorial no exige reproducir íntegramente la investigación. Su función es comprobar que el producto haya recorrido los controles correspondientes y que no subsista un riesgo que razonablemente impida su publicación.

3.5. Revisión especializada

La revisión ordinaria podrá complementarse con intervención especializada cuando la naturaleza del asunto lo requiera.

Revisión jurídica. Procederá cuando el contenido pueda generar consecuencias legales relevantes, particularmente en materias relacionadas con derechos de terceros, imputaciones de conductas ilícitas, procesos judiciales, privacidad, protección de datos, propiedad intelectual, confidencialidad o responsabilidad de la organización.

Revisión de protección de datos. Procederá cuando el tratamiento involucre categorías sensibles de datos, procesamiento a gran escala, perfilamiento, decisiones automatizadas u otras circunstancias que puedan representar un riesgo elevado para los titulares.

Revisión de seguridad. Será necesaria cuando se incorporen nuevas herramientas o integraciones, se procesen grandes volúmenes de información confidencial o exista un riesgo relevante para los sistemas, cuentas o repositorios institucionales.

Revisión técnica o temática. Se utilizará cuando la validez de una conclusión dependa de conocimientos especializados que el equipo responsable no pueda verificar razonablemente por sí mismo.

Estas revisiones podrán apoyarse en herramientas de inteligencia artificial, pero la asistencia tecnológica no sustituirá la intervención del especialista cuando esta resulte necesaria.

3.6. Administración institucional de las herramientas

Fundamedios mantendrá un registro de las herramientas de inteligencia artificial aprobadas para usos institucionales y de las restricciones que correspondan a cada una.

La autorización deberá considerar cinco dimensiones:

I. Funcionalidad: qué hace la herramienta y para qué usos institucionales se pretende utilizar.

II. Información: qué clases de datos podrá recibir y cuáles deberán mantenerse fuera de ella.

III. Privacidad: cómo utiliza, conserva y eventualmente reutiliza la información introducida, incluyendo su posible empleo para entrenamiento o mejora del servicio.

IV. Seguridad: qué mecanismos ofrece para autenticación, control de acceso, eliminación, protección y administración de información.

V. Condiciones jurídicas: qué terceros intervienen en el procesamiento, dónde puede procesarse la información y cuáles son las condiciones contractuales relevantes para Fundamedios.

Una herramienta podrá encontrarse autorizada para determinadas actividades y restringida para otras. La aprobación para corregir, traducir o estructurar información pública, por ejemplo, no implica autorización para introducir en ella documentos confidenciales o datos sensibles.

3.7. Escalamiento de riesgos

Cuando una persona encuentre durante el uso de IA una situación que no pueda resolver razonablemente dentro de sus funciones, deberá detener la parte afectada del proceso y solicitar la revisión correspondiente.

El escalamiento será obligatorio, entre otros supuestos, cuando exista una duda relevante sobre la autenticidad o veracidad de información esencial; aparezca inesperadamente información confidencial o especialmente sensible; pueda quedar comprometida una fuente periodística; exista un riesgo significativo para una persona; se genere una acusación que no encuentre sustento suficiente en las fuentes; se detecte un posible incidente de protección de datos o seguridad; o existan dudas razonables sobre la legalidad de un determinado tratamiento.

También deberá recurrirse a revisión especializada cuando el sistema produzca resultados cuya validación exija conocimientos técnicos que el usuario no posea.

Solicitar una revisión adicional frente a una situación de incertidumbre constituye un mecanismo normal de control y no un incumplimiento del usuario.

3.8. Separación de funciones según el riesgo

La intensidad de la supervisión dependerá del nivel de riesgo definido en la sección anterior.

En los **usos de riesgo bajo**, el propio usuario podrá operar la herramienta, revisar su resultado e incorporarlo a su trabajo dentro del flujo ordinario.

En los **usos de riesgo medio**, deberá existir una revisión humana efectiva de la información relevante antes de considerar terminado el producto.

Para los **usos de riesgo alto**, se exigirá una revisión reforzada y, siempre que resulte razonablemente posible, una segunda intervención humana sobre los hechos y conclusiones centrales. Si la naturaleza del riesgo lo exige, deberá añadirse la revisión jurídica, de protección de datos, de seguridad o técnica que corresponda.

Los **usos de riesgo crítico** no podrán continuar mientras subsista la circunstancia que determinó esa clasificación.

3.9. Determinación de responsabilidades

Cuando se produzca un incidente relacionado con el uso de inteligencia artificial, la existencia del error no determinará por sí sola la responsabilidad de quien operó la herramienta ni excluirá automáticamente la responsabilidad de quienes intervinieron posteriormente.

La revisión deberá reconstruir la cadena del proceso:

herramienta utilizada → información proporcionada → resultado generado → actuación del operador → verificación realizada → revisión → aprobación → publicación.

A partir de esa reconstrucción deberá determinarse si cada persona cumplió las funciones y controles que le correspondían.

Cuando el usuario haya empleado una herramienta autorizada, haya respetado las restricciones aplicables, haya documentado adecuadamente su utilización, haya efectuado las verificaciones exigidas y haya comunicado las limitaciones o dudas detectadas, la aparición posterior de un error no constituirá, por sí sola, un incumplimiento individual.

En cambio, podrán ser considerados incumplimientos, según su gravedad, la omisión de controles obligatorios; el uso deliberado de herramientas no autorizadas; la introducción indebida de información protegida; la utilización consciente de información no verificada; la eliminación u ocultamiento de registros que debían conservarse; o la ocultación de una intervención sustancial de inteligencia artificial cuando esta debía ser informada.

3.10. Responsabilidad de Fundamedios

Fundamedios deberá proporcionar las condiciones necesarias para la aplicación razonable de este Manual. Esto comprende la definición de procedimientos claros, la disponibilidad de herramientas adecuadas, mecanismos de consulta y escalamiento, capacitación periódica y medidas de seguridad acordes con los usos autorizados.

La gobernanza de inteligencia artificial se entenderá como una responsabilidad compartida: **la organización establece las condiciones y controles; cada persona responde por el cumplimiento de las funciones que le han sido asignadas dentro del proceso.**

El sistema deberá orientarse no solamente a identificar responsabilidades después de un incidente, sino principalmente a prevenirlo, detectarlo oportunamente, corregir sus efectos y evitar su repetición.

---

4. Reglas generales para la utilización de herramientas de inteligencia artificial

Las herramientas de inteligencia artificial utilizadas para actividades de Fundamedios deberán incorporarse al trabajo institucional de forma controlada. La disponibilidad pública de una plataforma, su gratuidad o su utilización habitual para fines personales no implica que pueda emplearse libremente para procesar información institucional.

Antes de utilizar una herramienta deberán considerarse tanto la tarea que se pretende realizar como la naturaleza de la información que será proporcionada al sistema.

4.1. Herramientas autorizadas

Fundamedios mantendrá un registro actualizado de las herramientas de inteligencia artificial autorizadas para uso institucional. El registro podrá establecer condiciones diferentes de utilización según las características de cada plataforma.

Las herramientas podrán clasificarse en tres categorías:

Herramientas de uso general. Podrán utilizarse para las actividades institucionales ordinarias expresamente permitidas, respetando las reglas de este Manual.

Herramientas de uso restringido. Podrán emplearse únicamente para determinadas actividades, clases de información, proyectos o usuarios.

Herramientas no autorizadas. No podrán utilizarse para procesar información institucional cuando sus condiciones de privacidad, seguridad, funcionamiento o tratamiento de datos no ofrezcan garantías suficientes, o cuando no hayan sido evaluadas para el uso que se pretende realizar.

La clasificación de una herramienta podrá modificarse cuando cambien sus funcionalidades, términos de servicio, políticas de privacidad, condiciones de seguridad o cualquier otra circunstancia relevante.

4.2. Evaluación previa de nuevas herramientas

La incorporación de una nueva herramienta de inteligencia artificial para usos institucionales relevantes deberá estar precedida por una evaluación proporcional a los riesgos que presente.

No será suficiente valorar únicamente la calidad de sus resultados. La evaluación deberá considerar también qué ocurre con la información que recibe.

En particular, deberá determinarse:

1. Finalidad. Qué función cumple la herramienta y qué necesidad institucional pretende resolver.

2. Datos de entrada. Qué información necesita recibir para funcionar y si resulta necesario introducir datos personales o información confidencial.

3. Conservación. Durante cuánto tiempo mantiene el proveedor los prompts, archivos, resultados u otra información asociada.

4. Reutilización. Si la información proporcionada puede utilizarse para entrenar, mejorar o desarrollar modelos o servicios del proveedor o de terceros.

5. Acceso. Quién puede acceder a la información y qué controles administrativos ofrece la plataforma.

6. Procesamiento por terceros. Si intervienen otros proveedores, subencargados o servicios tecnológicos.

7. Localización y transferencia. Cuando corresponda, dónde puede procesarse la información y qué implicaciones genera para la protección de datos personales.

8. Eliminación. Si Fundamedios puede eliminar conversaciones, archivos, registros u otros datos almacenados.

9. Seguridad. Qué mecanismos de autenticación, cifrado, control de acceso, registro de actividad y administración ofrece.

10. Condiciones jurídicas. Qué derechos y obligaciones establecen los términos contractuales respecto de la información introducida y de los resultados generados.

Una herramienta que no permita determinar razonablemente estas condiciones no deberá utilizarse para información sensible, confidencial o de riesgo alto.

4.3. Entorno institucional de inteligencia artificial y cuentas de usuario

Las actividades de Fundamedios que involucren inteligencia artificial deberán realizarse, como regla general, a través del **entorno institucional de inteligencia artificial autorizado por la organización**.

Este entorno estará compuesto por las cuentas, perfiles, espacios de trabajo, proyectos, asistentes, repositorios, configuraciones y demás herramientas que Fundamedios habilite para el desarrollo de sus actividades.

Su utilización tendrá una doble finalidad: **proteger la información institucional y preservar el conocimiento generado por la organización.**

Uso del entorno institucional

Cuando Fundamedios haya habilitado una herramienta institucional para una determinada actividad, el personal y los colaboradores deberán utilizarla en lugar de sus cuentas personales.

No deberán trasladarse investigaciones, documentos, bases de datos, prompts, conversaciones, resultados u otra información institucional a cuentas personales con el propósito de realizar actividades que puedan ejecutarse dentro del entorno institucional.

Tampoco deberán utilizarse cuentas personales para eludir restricciones de acceso, privacidad, seguridad, conservación de información o cualquier otro control establecido por Fundamedios.

Las excepciones deberán responder a una necesidad justificada y sujetarse a la autorización correspondiente.

Perfil institucional y conocimiento acumulado

Fundamedios procurará mantener un **perfil o entorno propio de inteligencia artificial**, configurado de acuerdo con las características, metodologías, actividades y necesidades de la organización.

Dentro de las capacidades que permitan las herramientas utilizadas, este entorno podrá incorporar de manera controlada información institucional destinada a mejorar la calidad y consistencia del trabajo, como metodologías,

manuales, criterios editoriales, formatos, documentos de referencia, terminología, antecedentes de proyectos y otros materiales previamente autorizados.

El objetivo será que el conocimiento generado mediante el uso institucional de inteligencia artificial pueda permanecer disponible para la organización y contribuir a trabajos posteriores, en lugar de quedar disperso entre cuentas individuales.

La incorporación de información a este entorno deberá respetar en todo momento las reglas de protección de datos, confidencialidad, propiedad intelectual, seguridad y conservación establecidas en este Manual.

La posibilidad técnica de incorporar información al conocimiento o memoria de una herramienta no constituye, por sí sola, autorización para hacerlo.

Titularidad y continuidad de la información institucional

Las investigaciones, instrucciones, configuraciones, bases documentales, prompts institucionales, resultados, metodologías y demás información producida o incorporada en el desarrollo de actividades para Fundamedios deberán conservarse, cuando corresponda, dentro de los sistemas y repositorios institucionales definidos para ese propósito.

La organización deberá evitar que información necesaria para continuar una investigación, reconstruir un proceso o aprovechar conocimiento previamente generado dependa exclusivamente de una cuenta personal de quien realizó el trabajo.

Cuando una persona deje de participar en un proyecto o termine su relación con Fundamedios, la información institucional que deba conservarse deberá permanecer accesible para la organización de acuerdo con las reglas de acceso, conservación y eliminación aplicables.

Esto no implica que toda conversación o interacción realizada con inteligencia artificial deba conservarse indefinidamente. El Manual establecerá posteriormente qué información debe formar parte del expediente institucional y durante cuánto tiempo deberá mantenerse.

Acceso individual dentro del entorno institucional

Siempre que la plataforma lo permita, cada usuario deberá acceder al entorno institucional mediante una identificación individual.

Los permisos deberán asignarse de acuerdo con las funciones y necesidades de cada persona. El acceso a un entorno institucional de IA no implicará automáticamente acceso a todas las investigaciones, proyectos, bases documentales o información disponible en la organización.

Las credenciales personales de acceso no deberán compartirse. Cuando una actividad requiera colaboración entre varias personas, deberán utilizarse las funciones de equipos, proyectos, espacios compartidos o mecanismos equivalentes proporcionados por la herramienta.

Este esquema permitirá conservar simultáneamente **la continuidad del conocimiento institucional y la trazabilidad individual de las actuaciones realizadas**.

Separación entre uso institucional y uso personal

Las cuentas personales de inteligencia artificial podrán utilizarse para actividades particulares de cada persona, pero no deberán convertirse en repositorios paralelos de información de Fundamedios.

Cuando una actividad se realice por cuenta de Fundamedios y genere información que deba conservarse, verificarse, compartirse o utilizarse posteriormente por la organización, deberá desarrollarse o incorporarse al entorno institucional correspondiente.

La separación entre cuentas personales e institucionales busca evitar que documentos, investigaciones, antecedentes, prompts, fuentes o conocimiento producido para Fundamedios queden almacenados fuera del control de la organización.

Configuración y administración

Las cuentas y entornos institucionales deberán configurarse utilizando, cuando se encuentren disponibles, las opciones de privacidad y seguridad más adecuadas para el tipo de información procesada.

Fundamedios determinará, según las características de cada herramienta, las configuraciones aplicables respecto de memoria, historial, conservación de conversaciones, utilización de información para entrenamiento o mejora de modelos, integraciones externas, acceso a archivos, permisos de usuarios y eliminación de información.

Estas configuraciones deberán revisarse periódicamente y cuando el proveedor introduzca cambios relevantes en el funcionamiento o en las condiciones de la herramienta.

4.4. Información que puede proporcionarse a una herramienta de IA

Antes de introducir información en un sistema, el usuario deberá identificar su naturaleza y preguntarse si la herramienta se encuentra autorizada para recibirla.

Como regla general, podrá utilizarse información pública, institucional no reservada o información previamente preparada para ese tratamiento, siempre que su utilización sea compatible con la finalidad correspondiente.

Cuando se trate de información que contenga datos personales, deberá aplicarse además el principio de minimización: se proporcionará únicamente aquella información necesaria para realizar la tarea.

Cuando resulte posible obtener el mismo resultado eliminando nombres, identificadores, datos de contacto u otros elementos innecesarios, deberá utilizarse la versión reducida, anonimizada o seudonimizada.

4.5. Información de acceso restringido

Determinadas categorías de información requerirán una autorización específica o una herramienta que haya sido aprobada para su tratamiento.

Esta regla comprenderá especialmente información relacionada con investigaciones no publicadas; documentos internos; bases de datos institucionales; información personal no pública; datos sensibles; expedientes jurídicos; información recibida bajo compromisos de confidencialidad; y documentación cuyo acceso se encuentre limitado dentro de la organización.

Antes de procesar este tipo de información mediante IA deberá comprobarse que la herramienta, cuenta y modalidad de utilización ofrecen las garantías necesarias.

Cuando exista duda, la información no deberá introducirse hasta realizar la consulta correspondiente.

4.6. Información que no deberá introducirse en herramientas de uso general

Con independencia de la utilidad que pueda ofrecer una plataforma, no deberán incorporarse en herramientas de IA de uso general:

a) contraseñas, claves privadas, códigos de recuperación, tokens o credenciales de acceso;

b) información que permita el acceso no autorizado a sistemas, cuentas o infraestructura;

c) identidad de fuentes periodísticas confidenciales o información que permita razonablemente deducirla;

d) información cuya revelación pueda poner en peligro la integridad física o seguridad de una persona;

e) documentación sometida a reserva legal cuando la herramienta no se encuentre expresamente habilitada para su tratamiento; y

f) cualquier información respecto de la cual Fundamedios haya asumido una obligación de confidencialidad incompatible con su procesamiento por el proveedor de la herramienta.

Esta prohibición comprende tanto la introducción directa de la información en un prompt como su incorporación mediante archivos, imágenes, enlaces privados, bases de datos, integraciones u otros mecanismos.

4.7. Documentos y archivos adjuntos

Cargar un archivo en una herramienta de inteligencia artificial constituye una forma de proporcionar su contenido al sistema.

Antes de adjuntar un documento deberá revisarse si contiene información que no resulte necesaria para la tarea.

Cuando corresponda, deberá prepararse previamente una copia destinada al procesamiento mediante IA, eliminando o protegiendo la información que no deba ser compartida.

Deberá prestarse especial atención a información que pueda encontrarse oculta o pasar inadvertida en documentos extensos, anexos, hojas de cálculo, historiales, comentarios, metadatos o conjuntos de documentos.

El hecho de que la herramienta permita técnicamente cargar un archivo no significa que su contenido pueda ser procesado legítimamente.

4.8. Uso de información obtenida desde Internet y redes sociales

La posibilidad de que una herramienta de inteligencia artificial acceda, encuentre o procese información disponible en Internet no convierte automáticamente esa información en verdadera, actualizada o legítimamente reutilizable.

Cuando la IA sea utilizada para búsquedas abiertas, monitoreo de medios o redes sociales, identificación de publicaciones o recopilación de antecedentes, deberán conservarse las referencias necesarias para regresar a la fuente original.

Siempre que un dato resulte relevante para una investigación o publicación, deberá verificarse directamente su origen y contexto.

Una síntesis generada por IA sobre una página web, publicación, cuenta de redes sociales o documento no sustituirá la consulta de la fuente cuando de ella dependa una afirmación relevante.

4.9. Uso de información generada por la propia IA

Deberá mantenerse una distinción clara entre:

información recuperada, que proviene de una fuente externa identificable;

información transformada, que ha sido resumida, traducida, clasificada o reorganizada por la herramienta; e

información generada, que ha sido producida por el modelo a partir de sus propios procesos de inferencia.

Esta distinción deberá considerarse durante todo el proceso.

Cuanto mayor sea la intervención generativa de la herramienta, mayor deberá ser la precaución antes de considerar su resultado como sustento factual.

La capacidad de una IA para proporcionar una cita, enlace, referencia bibliográfica o supuesto documento tampoco garantiza su existencia o correspondencia. Las referencias relevantes deberán comprobarse en su fuente original.

4.10. Prohibición de eludir controles

Ninguna persona deberá utilizar una herramienta, cuenta personal, servicio externo o mecanismo alternativo con el propósito de evitar las restricciones establecidas en este Manual.

Cuando una herramienta autorizada no permita realizar determinada actividad debido a sus controles de seguridad, privacidad o acceso, esta circunstancia no habilitará automáticamente a realizarla mediante otra plataforma.

Las necesidades que no puedan satisfacerse mediante las herramientas disponibles deberán comunicarse para evaluar una solución adecuada.

4.11. Cambios en las herramientas

Las herramientas de inteligencia artificial evolucionan continuamente. Una misma plataforma puede modificar sus modelos, capacidades, acceso a Internet, mecanismos de memoria, políticas de conservación, integraciones, términos contractuales o utilización de los datos.

Por esta razón, una autorización institucional no tendrá carácter indefinido.

Las herramientas utilizadas de manera habitual deberán revisarse periódicamente y también cuando se produzca un cambio significativo que pueda afectar su seguridad, privacidad, funcionamiento o condiciones de utilización.

Si se detecta un cambio que pueda incrementar sustancialmente el riesgo, Fundamedios podrá restringir o suspender temporalmente el uso de la herramienta hasta completar su evaluación.

4.12. Preguntas Generales

Ante cualquier utilización de inteligencia artificial deberán poder responderse, antes de continuar, cinco preguntas básicas:

¿Para qué se utilizará? La finalidad debe encontrarse claramente definida.

¿Qué información recibirá? Debe conocerse su naturaleza, sensibilidad y necesidad.

¿La herramienta puede recibir esa información? Debe encontrarse autorizada para el uso y nivel de información correspondientes.

¿Qué riesgo genera el resultado? Deberá aplicarse el nivel de control definido en este Manual.

¿Quién verificará el resultado? Toda utilización sustantiva deberá tener una persona responsable de la supervisión correspondiente.

Si alguna de estas preguntas no puede responderse razonablemente, el proceso deberá detenerse hasta contar con la información o autorización necesaria.

CAPÍTULO

II

Investigación y levantamiento de información mediante inteligencia artificial



La inteligencia artificial puede ampliar considerablemente la capacidad de Fundamedios para investigar hechos, localizar documentos, monitorear información pública, procesar grandes cantidades de contenido e identificar relaciones que posteriormente puedan ser examinadas por periodistas, investigadores, verificadores y analistas.

Su utilidad en esta etapa radica principalmente en **facilitar el acceso y procesamiento de información**, no en sustituir las fuentes que sustentan una investigación.

Por esta razón, el levantamiento de información mediante IA deberá mantener una separación clara entre aquello que el sistema **encuentra en una fuente**, aquello que **procesa o interpreta a partir de una fuente** y aquello que **genera por inferencia propia**. Esta distinción acompañará todo el procedimiento desarrollado en este capítulo.

El ciclo de investigación asistida por IA comprende cuatro momentos principales:

planificar la búsqueda → recolectar información y fuentes → verificar los hallazgos → organizar y validar la información

Solo después de cumplir los controles que correspondan podrá la información obtenida mediante este proceso utilizarse como insumo para la elaboración de contenidos.

1. Planificación de la investigación

No toda consulta realizada a una herramienta de inteligencia artificial constituye una investigación. Una pregunta exploratoria puede utilizarse para comprender inicialmente un asunto, identificar conceptos o descubrir posibles líneas de búsqueda sin necesidad de activar un procedimiento formal.

Cuando la IA vaya a utilizarse para obtener información que pueda sustentar una alerta, verificación, investigación, nota, informe u otro producto de Fundamedios, deberá existir, en cambio, una **finalidad de investigación suficientemente definida**.

1.1. Definir qué se busca

Antes de iniciar una investigación sustantiva deberá identificarse, al menos de manera general, qué se pretende conocer.

La pregunta inicial puede ser amplia. En ocasiones precisamente se utilizará la inteligencia artificial porque todavía no se conocen todos los elementos del asunto. Sin embargo, el investigador deberá poder distinguir entre una búsqueda abierta destinada a descubrir información y una búsqueda dirigida a comprobar un hecho específico.

Por ejemplo, no es lo mismo utilizar IA para conocer inicialmente los antecedentes de una controversia que utilizarla para determinar si una persona ocupó determinado cargo, recibió un contrato, realizó una declaración o participó en un acontecimiento.

A medida que la investigación avance, las preguntas deberán hacerse progresivamente más específicas.

Una investigación puede comenzar preguntando:

¿Qué se conoce sobre este asunto?

y posteriormente avanzar hacia:

¿Qué ocurrió? ¿Cuándo ocurrió? ¿Quién intervino? ¿Qué documento lo demuestra? ¿Cuál es la fuente original? ¿Existen versiones diferentes? ¿Qué información contradice el hallazgo inicial?

El propósito es pasar gradualmente de la **exploración** a la **comprobación**.

1.2. Determinar para qué se utilizará la información

La finalidad prevista condicionará la profundidad de la investigación y el nivel de verificación requerido.

Una búsqueda destinada a preparar una reunión interna no necesariamente requiere los mismos controles que una investigación cuyo resultado identificará públicamente a una persona como responsable de determinada conducta.

Antes de comenzar deberá conocerse, cuando sea posible, si la información servirá para una alerta, una verificación de Ecuador Chequea, un monitoreo de Liberties 360°, una investigación periodística, una publicación institucional, un informe, un análisis interno u otro producto.

La finalidad podrá cambiar durante el trabajo. Si una búsqueda inicialmente exploratoria evoluciona hacia una investigación destinada a publicación, deberán aplicarse desde ese momento los controles correspondientes a su nueva finalidad y nivel de riesgo.

1.3. Delimitar la búsqueda cuando sea necesario

Una instrucción excesivamente abierta puede provocar que la herramienta mezcle períodos, jurisdicciones, personas con nombres similares, versiones desactualizadas o acontecimientos distintos.

Cuando estos elementos sean relevantes, la búsqueda deberá delimitarse mediante referencias como:

- período o fecha;
- país, ciudad o jurisdicción;
- nombre completo de las personas u organizaciones involucradas;
- cargo o función;
- materia investigada;
- tipo de documento buscado;
- fuentes que deben priorizarse;
- idioma, cuando pueda afectar los resultados; o
- cualquier otra característica que permita distinguir el objeto de investigación.

No todos estos elementos deberán aparecer en cada prompt. Deberán utilizarse aquellos que ayuden efectivamente a reducir ambigüedades.

1.4. Establecer una estrategia de fuentes

La inteligencia artificial deberá utilizarse, siempre que sea posible, como un mecanismo para **acercar al investigador a la evidencia original**.

La estrategia de fuentes dependerá del objeto de la investigación. No existe una jerarquía absoluta que funcione para todos los casos: una declaración publicada directamente por una persona puede ser la fuente primaria para demostrar qué dijo; un Registro Oficial puede ser la fuente adecuada para comprobar un nombramiento; un expediente judicial para conocer una decisión; y una entrevista directa puede aportar información que todavía no existe en ningún documento público.

Sin perjuicio de ello, cuando se investiguen hechos susceptibles de comprobación documental, deberán priorizarse las fuentes que permitan una mayor trazabilidad y verificación.

Podrán considerarse, entre otras:

I. Fuentes primarias u originales: documentos oficiales, registros públicos, expedientes, bases institucionales, contratos, resoluciones, sentencias, entrevistas, declaraciones originales, grabaciones y publicaciones realizadas directamente por la persona u organización correspondiente.

II. Fuentes secundarias calificadas: investigaciones periodísticas, publicaciones académicas, informes de organizaciones especializadas, organismos internacionales y otros materiales que analicen o documenten información proveniente de fuentes identificables.

III. Fuentes abiertas complementarias: redes sociales, blogs, repositorios colaborativos, foros, publicaciones individuales y otras fuentes que puedan aportar indicios, antecedentes o elementos que posteriormente requieran comprobación.

Una respuesta producida por una IA **no constituye una cuarta categoría de fuente**. Es una herramienta para localizar, procesar o relacionar información que deberá poder vincularse posteriormente con una fuente identificable.

1.5. Investigar sin confirmar previamente una conclusión

Las instrucciones proporcionadas a la IA deberán formularse, especialmente en asuntos controvertidos, de manera que permitan encontrar información que confirme **y también que contradiga** la hipótesis inicial.

Debe evitarse orientar el sistema exclusivamente hacia la búsqueda de evidencia favorable a una conclusión previamente asumida.

Por ejemplo, ante una denuncia sobre una posible irregularidad, resultará preferible solicitar:

Identifica fuentes y documentos que permitan determinar si ocurrió la conducta señalada. Busca información que pueda confirmarla, descartarla o modificar su contexto. Distingue hechos comprobados, versiones de las partes e información que no haya podido verificarse.

antes que formular una instrucción cuyo punto de partida sea que la irregularidad efectivamente ocurrió.

Esta precaución es especialmente importante porque los sistemas generativos pueden producir respuestas que se acomodan a la premisa contenida en la pregunta, aun cuando esa premisa sea incompleta o incorrecta.

1.6. Definir qué deberá entregar la IA

En una búsqueda sustantiva no debería solicitarse únicamente una explicación narrativa. Siempre que la herramienta tenga capacidad para acceder a fuentes, la instrucción deberá procurar que los hallazgos puedan ser comprobados posteriormente.

Dependiendo del trabajo, podrá pedirse al sistema que entregue:

el hallazgo, expresado de forma concreta;

la fuente, identificando de dónde procede;

el enlace o referencia, cuando exista;

la fecha, cuando resulte relevante;

el fragmento o ubicación pertinente, sin sustituir la revisión del documento original;

el nivel de certeza, diferenciando lo comprobado de aquello que permanece dudoso; y

las contradicciones, cuando existan fuentes relevantes que presenten una versión diferente.

Cuando una herramienta afirme haber encontrado un documento o información relevante pero no pueda proporcionar una referencia que permita localizarlo, ese resultado deberá permanecer como **información pendiente de verificación**.

No deberá rellenarse esa ausencia suponiendo que la fuente existe.

1.7. Información sensible desde el inicio

La protección de la información comienza antes de realizar la primera búsqueda.

Si razonablemente puede anticiparse que una investigación involucrará fuentes confidenciales, víctimas, denunciantes, menores de edad, datos sensibles, información jurídica reservada, documentación no publicada o información cuya revelación pueda generar riesgos de seguridad, deberá definirse previamente qué información puede ser proporcionada a la herramienta.

En estos casos puede ser necesario sustituir nombres por identificadores, retirar información innecesaria de los documentos, trabajar con copias previamente depuradas o utilizar únicamente entornos institucionales específicamente autorizados para ese nivel de información.

Si para realizar una búsqueda no es necesario que la IA conozca la identidad de una persona, **no deberá proporcionársele esa identidad**.

1.8. Uso del entorno institucional durante la investigación

Las investigaciones realizadas para Fundamedios deberán desarrollarse dentro del entorno institucional de inteligencia artificial definido en este Manual cuando exista una herramienta habilitada para esa actividad.

Esto permite que las búsquedas relevantes, fuentes localizadas, metodologías empleadas y demás conocimiento que corresponda conservar permanezcan vinculados al trabajo de la organización y no dependan de las cuentas personales de quienes participan en la investigación.

Cuando varios miembros del equipo trabajen sobre un mismo asunto, deberán utilizarse, siempre que las capacidades de la plataforma lo permitan, proyectos, espacios compartidos o mecanismos institucionales que permitan preservar continuidad sin eliminar la trazabilidad individual.

La existencia de un entorno institucional común no significa que toda la información acumulada por Fundamedios deba ponerse automáticamente a disposición de cada usuario o de cada sistema de IA. El acceso continuará sujeto a las reglas de confidencialidad, necesidad y permisos correspondientes.

1.9. Cuándo termina la planificación

La planificación estará suficientemente cumplida cuando el investigador pueda responder razonablemente tres preguntas:

¿Qué estoy tratando de averiguar?

¿Qué tipo de evidencia necesito para comprobarlo?

¿Qué riesgos debo controlar mientras la busco?

No será necesario producir un formulario para responder estas preguntas en cada investigación ordinaria. En asuntos de riesgo alto, sin embargo, estas definiciones deberán poder reconstruirse posteriormente como parte de la trazabilidad del proceso.

La planificación debe facilitar la investigación, no retrasarla. Su función es evitar que la inteligencia artificial transforme una búsqueda dirigida en una acumulación indiscriminada de información sin propósito, sin fuentes identificables y sin criterios posteriores de verificación.

2. Búsqueda y recolección de información mediante inteligencia artificial

Una vez definido el objeto de investigación, la inteligencia artificial podrá utilizarse para localizar información, documentos, fuentes, antecedentes y relaciones relevantes para el trabajo de Fundamedios.

La búsqueda mediante IA no deberá entenderse como una consulta única. En investigaciones sustantivas será normalmente un **proceso progresivo**, en el cual los primeros resultados permiten formular nuevas preguntas, localizar fuentes más precisas, detectar vacíos y orientar las siguientes búsquedas.

Durante esta etapa deberá mantenerse una diferencia fundamental entre **descubrir información** y **comprobar información**. La IA puede ser particularmente eficiente para lo primero; lo segundo exige regresar a las fuentes.

2.1. La IA como herramienta de búsqueda y descubrimiento

La inteligencia artificial podrá utilizarse para ampliar las capacidades ordinarias de investigación. Entre otros usos, podrá ayudar a localizar documentos; identificar antecedentes; encontrar publicaciones anteriores; relacionar nombres, fechas o acontecimientos; construir cronologías preliminares; identificar instituciones que puedan poseer determinada información; descubrir términos alternativos de búsqueda; localizar posibles fuentes; comparar grandes cantidades de resultados; o señalar asuntos que merezcan una investigación adicional.

También podrá utilizarse para formular nuevas preguntas a partir de la información encontrada.

Este último uso es especialmente relevante. Una buena investigación asistida por IA no debería limitarse a preguntar al sistema por una conclusión, sino utilizarlo para **descubrir qué debería investigarse a continuación**.

Por ejemplo, a partir de un documento localizado podrá solicitarse que identifique nombres, fechas, referencias a otros documentos, instituciones involucradas o acontecimientos que puedan convertirse en nuevas líneas de búsqueda.

Cada una de esas líneas deberá posteriormente ser examinada de acuerdo con su relevancia y nivel de riesgo.

2.2. Búsquedas sucesivas

Cuando la primera consulta produzca información relevante, el investigador deberá profundizar progresivamente en los hallazgos en lugar de asumir como definitiva la respuesta inicial.

Una secuencia razonable puede avanzar de esta manera:

Exploración ¿Qué información existe sobre el asunto?

Identificación ¿Qué documentos, personas, instituciones y acontecimientos aparecen relacionados?

Profundización ¿Qué dicen exactamente esas fuentes?

Comprobación ¿Existe evidencia independiente que confirme los principales hechos?

Contraste ¿Existen documentos o fuentes que contradigan, modifiquen o contextualicen esos hallazgos?

Cierre ¿Qué hechos pueden considerarse suficientemente comprobados y cuáles permanecen abiertos?

No todas las investigaciones necesitarán recorrer estas etapas de manera formal ni en ese orden exacto. Constituyen una guía para evitar que una respuesta inicial de la herramienta se convierta prematuramente en la conclusión de la investigación.

2.3. Cada hallazgo debe conducir a una fuente

Cuando una herramienta de IA proporcione información factual relevante, el siguiente objetivo será determinar **de dónde proviene esa información**.

Si el sistema dispone de búsqueda o acceso a Internet, deberá solicitarse la fuente correspondiente. Cuando proporcione un enlace, este deberá ser abierto y revisado antes de considerar comprobado el dato.

No bastará con que el enlace exista. Será necesario comprobar que:

a) corresponde efectivamente a la fuente indicada;

b) contiene la información atribuida;

c) la información aparece en el contexto descrito;

d) la fecha corresponde al período investigado;

e) no se trata de una persona, institución o acontecimiento diferente; y

f) la fuente no ha sido citada de manera parcial o engañosa.

Una referencia aparentemente precisa puede ser incorrecta. Los sistemas de inteligencia artificial pueden generar títulos, autores, enlaces, números de procesos, normas, sentencias, publicaciones o citas inexistentes o combinar elementos provenientes de fuentes diferentes.

Por ello, **la precisión aparente de una referencia no reemplaza su comprobación**.

2.4. Cuando la IA no proporciona la fuente

Una situación frecuente ocurre cuando el sistema ofrece una afirmación concreta pero no puede identificar adecuadamente el documento, publicación o fuente que supuestamente la sustenta.

En ese caso, la información podrá conservarse temporalmente como **pista de investigación**, pero no deberá tratarse como un hecho verificado.

El investigador podrá intentar localizar la fuente mediante nuevas búsquedas, utilizando elementos específicos de la respuesta: nombres, fechas, frases, instituciones, números de documentos u otros datos que permitan rastrear su posible origen.

Si después de una búsqueda razonable la fuente no puede localizarse, la información no deberá incorporarse como hecho a un producto público.

La ausencia de una fuente no deberá resolverse solicitando reiteradamente a la misma herramienta que confirme su propia respuesta. **Una IA no verifica una afirmación simplemente repitiéndola.**

2.5. Fuentes primarias y regreso al documento original

Cuando una fuente secundaria haga referencia a un documento primario relevante, deberá procurarse localizar el documento original.

Si una nota periodística señala que una sentencia estableció determinada cuestión, deberá revisarse la sentencia cuando esté disponible. Si un informe menciona una cifra proveniente de una institución pública, deberá buscarse la estadística o base correspondiente. Si una publicación atribuye una declaración a una persona, deberá localizarse, cuando resulte posible, la entrevista, intervención, publicación o registro original.

Esta regla adquiere mayor importancia mientras más central sea el dato para la investigación.

No significa que las fuentes secundarias carezcan de valor. Pueden aportar contexto, investigaciones propias, testimonios y análisis que no aparecen en documentos oficiales. El objetivo es evitar que la IA multiplique cadenas de referencias en las que ninguna persona haya revisado finalmente el origen del dato.

2.6. Recolección desde medios de comunicación

Los medios de comunicación podrán utilizarse como fuentes de antecedentes y, cuando corresponda, como fuentes periodísticas de información.

Al utilizar IA para recopilar noticias deberá prestarse atención a la posibilidad de que el sistema mezcle informaciones publicadas en momentos distintos o presente como actual una situación que posteriormente cambió.

Por ello, deberán conservarse las fechas de publicación y distinguirse entre el momento en que ocurrió un hecho y el momento en que fue reportado.

Cuando varios medios reproduzcan la misma información, no deberá asumirse automáticamente que existen varias confirmaciones independientes. El investigador deberá determinar, cuando sea relevante, si las publicaciones se sustentan en fuentes distintas o si todas reproducen una misma agencia, comunicado, declaración o publicación original.

La repetición de un dato en numerosos sitios no equivale necesariamente a corroboración independiente.

2.7. Redes sociales y plataformas digitales

Las publicaciones en redes sociales pueden constituir tanto una fuente directa como una pista de investigación.

Una publicación realizada desde una cuenta auténtica puede ser fuente primaria para demostrar que determinada cuenta publicó cierto contenido. Sin embargo, ello no demuestra necesariamente que las afirmaciones contenidas en esa publicación sean verdaderas.

Cuando una publicación sea relevante deberán considerarse aspectos como su autoría, fecha, contexto, integridad y permanencia.

Particular atención deberá prestarse a:

- cuentas falsas o de suplantación;
- contenido republicado sin contexto;
- capturas de pantalla cuya fuente original no pueda encontrarse;
- publicaciones editadas o eliminadas;
- imágenes o videos reutilizados de acontecimientos anteriores;
- contenido generado o manipulado mediante inteligencia artificial; y
- cuentas cuyo nombre o apariencia puedan confundirse con fuentes oficiales.

Cuando resulte necesario preservar una publicación relevante para una investigación, deberá conservarse evidencia suficiente de su existencia y contexto de acuerdo con los procedimientos de documentación establecidos en este Manual.

2.8. Información audiovisual

La inteligencia artificial podrá utilizarse para transcribir, traducir, describir, clasificar o apoyar el análisis de fotografías, grabaciones, transmisiones, audios y videos.

Sus resultados deberán considerarse auxiliares.

Una transcripción automática puede omitir palabras; un sistema de reconocimiento puede confundir personas u objetos; una traducción puede modificar el sentido de una expresión; y una herramienta de análisis audiovisual puede realizar inferencias incorrectas.

Cuando una afirmación dependa de una parte específica de un audio o video, la persona responsable deberá revisar directamente el segmento correspondiente.

En materiales sensibles o controvertidos, la determinación de autenticidad no deberá descansar exclusivamente en la afirmación de una herramienta de IA. Podrá requerirse análisis adicional, búsqueda del archivo original, revisión de metadatos, comparación con otras fuentes o intervención técnica especializada.

2.9. Documentos extensos y bases documentales

Una de las principales ventajas de la inteligencia artificial es su capacidad para apoyar el análisis de grandes cantidades de documentos.

Podrá utilizarse para identificar menciones, extraer datos, organizar documentos por temas, comparar versiones, elaborar cronologías, localizar contradicciones o resumir conjuntos documentales.

Sin embargo, cuando un hallazgo vaya a utilizarse sustancialmente, deberá regresar al documento que lo contiene.

La regla será sencilla:

La IA puede indicar dónde mirar; el investigador debe comprobar qué dice realmente el documento.

Cuando la herramienta identifique una afirmación relevante dentro de un documento extenso, deberá procurarse conservar la página, sección, párrafo, registro u otra referencia que permita encontrarla nuevamente.

2.10. Citas y declaraciones

Las citas textuales requieren un estándar particularmente alto de comprobación.

Una frase entre comillas no deberá atribuirse a una persona únicamente porque una IA señale que la pronunció.

Antes de utilizar una cita deberá localizarse la fuente en la que pueda comprobarse razonablemente:

quién la dijo — qué dijo — cuándo — dónde — y en qué contexto.

Si únicamente puede comprobarse el sentido general de una declaración, pero no las palabras exactas, deberá utilizarse una paráfrasis correctamente atribuida y no una cita textual fabricada o reconstruida.

La IA podrá ayudar a localizar una declaración dentro de entrevistas, videos, documentos o transcripciones, pero la cita destinada a publicación deberá contrastarse con el material original siempre que este se encuentre disponible.

2.11. Datos, cifras y estadísticas

Cuando una investigación utilice cifras obtenidas mediante IA deberá identificarse la fuente de los datos, el período al que corresponden y, cuando resulte relevante, la metodología utilizada para producirlos.

Deberá evitarse combinar cifras provenientes de años, universos o metodologías diferentes como si fueran directamente comparables.

La IA podrá utilizarse para realizar cálculos, ordenar información o detectar tendencias, pero los resultados cuantitativos relevantes deberán revisarse antes de su utilización.

En particular, cuando el sistema realice cálculos sobre una base proporcionada por Fundamedios, deberán conservarse los datos de origen y la metodología suficiente para reproducir o revisar el resultado cuando la importancia de la conclusión así lo justifique.

2.12. Hallazgos inesperados

Durante una investigación pueden aparecer datos que no formaban parte de la finalidad inicial.

Un hallazgo inesperado no deberá utilizarse automáticamente.

Si la nueva información resulta relevante, deberá evaluarse si corresponde ampliar formalmente la investigación, qué riesgos adicionales genera y si su tratamiento resulta compatible con las reglas de protección de datos y confidencialidad.

Cuando el hallazgo involucre información especialmente sensible, una fuente protegida, una posible amenaza o datos cuya divulgación pueda producir un daño significativo, deberá suspenderse su procesamiento ordinario y aplicarse el mecanismo de escalamiento correspondiente.

2.13. No recopilar por recopilar

La capacidad técnica de los sistemas de inteligencia artificial para procesar grandes cantidades de información no justifica construir archivos indiscriminados sobre personas o instituciones.

La recolección deberá mantener relación con una finalidad periodística, investigativa, de verificación, monitoreo o institucional identificable.

Especialmente cuando existan datos personales, deberá evitarse conservar información que no aporte razonablemente al objeto de la investigación.

La pregunta no deberá ser únicamente **“¿podemos obtener este dato?”**, sino también **“¿necesitamos este dato para este trabajo?”**.

2.14. Resultado de la etapa de recolección

Al terminar la búsqueda, la información obtenida no deberá considerarse automáticamente lista para ser utilizada en una publicación.

El resultado de esta etapa será un conjunto organizado de:

fuentes localizadas, documentos obtenidos, datos recopilados, posibles hallazgos, contradicciones identificadas y asuntos todavía pendientes de comprobación.

A partir de ese momento comienza una etapa diferente: la **verificación y validación de la información**.

Hasta completar esa etapa deberá mantenerse claramente identificada la información que todavía sea provisional, inferida, contradictoria o no verificable.

3. Verificación y validación de la información obtenida mediante inteligencia artificial

La verificación constituye la etapa mediante la cual Fundamedios determina si la información obtenida, localizada, procesada o sugerida mediante inteligencia artificial cuenta con respaldo suficiente para ser utilizada posteriormente en una investigación, análisis, alerta, verificación, nota periodística, informe u otro producto institucional.

La inteligencia artificial puede producir respuestas convincentes, ordenadas y aparentemente bien sustentadas incluso cuando contienen errores, omisiones, referencias inexistentes, información desactualizada, interpretaciones incorrectas o inferencias que no se desprenden de las fuentes originales. Por esta razón, la apariencia de precisión de una respuesta no podrá sustituir el proceso de comprobación.

La validación no consiste únicamente en confirmar que una fuente exista. También requiere verificar que la fuente efectivamente diga lo que se le atribuye, que corresponda al momento y contexto investigados, que no haya sido citada de forma parcial o engañosa, que la información no haya sido modificada posteriormente y que el nivel de certeza con el cual se presenta el hallazgo sea compatible con la evidencia disponible.

El principio aplicable será:

La inteligencia artificial puede señalar un posible hecho; la validación determina si ese hecho puede ser utilizado por Fundamedios.

3.1. Estados de la información durante la investigación

Toda información relevante obtenida mediante inteligencia artificial deberá poder ubicarse, al menos conceptualmente, dentro de uno de los siguientes estados:

Información exploratoria. Es aquella que aparece durante una búsqueda inicial y sirve para orientar nuevas líneas de investigación. No requiere todavía un nivel elevado de certeza y no deberá utilizarse como afirmación factual en un producto público.

Información pendiente de verificación. Es aquella que presenta suficiente relevancia para ser investigada, pero cuya fuente, exactitud, contexto o vigencia todavía no han sido comprobados de manera satisfactoria.

Información parcialmente verificada. Es aquella respecto de la cual existe evidencia de respaldo, pero subsisten elementos importantes pendientes de confirmación, contradicciones entre fuentes o limitaciones que impiden considerarla plenamente acreditada.

Información verificada. Es aquella cuya fuente, contenido, contexto y relevancia han sido comprobados de manera suficiente para la finalidad específica de la investigación.

Información validada para uso editorial. Es aquella que, además de encontrarse verificada, ha cumplido los controles adicionales exigidos por su nivel de riesgo y puede ser utilizada como insumo en la producción de un contenido.

Esta clasificación no exige necesariamente utilizar etiquetas formales en cada investigación. Su finalidad es evitar que la información pase inadvertidamente de una respuesta preliminar de IA a una publicación sin haber atravesado los controles necesarios.

3.2. Qué significa verificar un dato

Verificar una información no significa simplemente encontrar otra página que repita la misma afirmación.

La comprobación deberá considerar, según corresponda, al menos los siguientes elementos:

a) Existencia de la fuente. Deberá comprobarse que el documento, publicación, registro, declaración, base de datos o material citado realmente exista y pueda ser localizado de manera razonable.

b) Correspondencia. La fuente deberá contener efectivamente la información que se pretende atribuirle. No será suficiente que trate del mismo asunto de forma general.

c) Autoría. Cuando resulte relevante, deberá comprobarse quién produjo, publicó o emitió la fuente y si puede atribuirse razonablemente a la persona o institución correspondiente.

d) Fecha. La información deberá corresponder al período investigado. Una fuente auténtica pero desactualizada puede producir una conclusión incorrecta.

e) Contexto. Deberá analizarse el contenido suficiente para evitar que una frase, cifra o afirmación cambie de significado al ser separada del contexto en el que aparece.

f) Integridad. Cuando existan indicios de edición, fragmentación, manipulación o alteración, deberá buscarse una versión más completa u original antes de utilizar el contenido.

g) Vigencia. En materias jurídicas, institucionales, estadísticas o regulatorias deberá comprobarse si la información continúa vigente o ha sido modificada posteriormente.

h) Pertinencia. La fuente puede ser auténtica y correcta, pero no necesariamente adecuada para demostrar la afirmación concreta que se pretende realizar.

La intensidad de esta comprobación dependerá de la importancia del dato dentro del producto final.

3.3. Verificación de hechos centrales y hechos accesorios

No todos los elementos de una investigación tienen la misma importancia.

Los **hechos centrales** son aquellos de los cuales depende sustancialmente el sentido, conclusión o impacto de una publicación. Si ese hecho fuera incorrecto, el producto podría resultar engañoso, injustificado o sustancialmente diferente.

Los **hechos accesorios** aportan contexto, antecedentes o información complementaria, pero su eventual modificación no cambia la conclusión principal.

Los hechos centrales deberán recibir el nivel más alto de verificación razonablemente posible.

Cuando una publicación atribuya una irregularidad, responsabilidad, conducta, declaración, vínculo, participación o antecedente relevante a una persona o institución, la información que sustente esa atribución será considerada central.

La revisión de un producto no deberá limitarse a determinar si “en términos generales” la historia parece correcta. Deberá identificar cuáles son las afirmaciones sobre las cuales realmente descansa el contenido y comprobarlas individualmente.

3.4. Corroboración independiente

Cuando la naturaleza y riesgo de una afirmación lo requieran, deberá procurarse corroboración mediante más de una fuente.

Sin embargo, el número de publicaciones que repiten una afirmación no constituye por sí mismo evidencia de corroboración independiente.

Si diez medios reproducen una misma declaración proveniente de una agencia, un comunicado o una publicación en redes sociales, puede existir una sola fuente original.

Por ello, cuando se utilice la existencia de varias fuentes como elemento de validación, deberá examinarse si estas tienen **orígenes realmente independientes**.

La corroboración podrá producirse, por ejemplo, mediante:

- un documento oficial y una entrevista directa;
- un registro público y una fuente periodística independiente;
- dos documentos producidos por instituciones diferentes;
- evidencia documental y material audiovisual;
- declaraciones independientes de personas con conocimiento directo; o
- diferentes conjuntos de datos que permitan confirmar una misma conclusión.

La cantidad necesaria de fuentes dependerá del nivel de riesgo. En determinadas investigaciones una única fuente primaria puede ser suficiente para demostrar un hecho objetivo; en otras, incluso varias fuentes pueden resultar insuficientes si dependen unas de otras o presentan dudas de credibilidad.

3.5. Regla reforzada para afirmaciones que puedan afectar a personas

Cuando la información pueda afectar significativamente la honra, reputación, privacidad, seguridad, posición jurídica o intereses de una persona identificada o identificable, la verificación deberá ser reforzada.

Antes de utilizar este tipo de información deberán analizarse de manera particularmente cuidadosa:

Primero. La calidad y origen de la evidencia.

Segundo. La posibilidad de que la información haya sido interpretada incorrectamente por la IA.

Tercero. La existencia de documentos o versiones que contradigan la conclusión inicial.

Cuarto. El contexto necesario para evitar una presentación incompleta de los hechos.

Quinto. La pertinencia de realizar contraste con la persona o institución involucrada, de acuerdo con los estándares periodísticos aplicables.

Una IA no deberá ser utilizada para “completar” una acusación cuando existan vacíos en la evidencia.

Cuando la información disponible únicamente permita formular una hipótesis, deberá mantenerse como hipótesis y no convertirse lingüísticamente en una afirmación categórica durante la generación de contenidos.

3.6. Verificación de citas textuales

Las citas presentan un riesgo particular porque los sistemas generativos pueden reconstruir frases plausibles que nunca fueron pronunciadas literalmente.

Por ello, una cita textual deberá verificarse directamente en el material que permita conocer las palabras efectivamente utilizadas.

Cuando sea posible, deberán comprobarse:

quién habló; qué dijo exactamente; cuándo lo dijo; dónde lo dijo; a qué pregunta o circunstancia respondió; y si la cita ha sido recortada de manera que altere su sentido.

Una cita obtenida de una transcripción automática deberá contrastarse con el audio o video original cuando sea sustancial para la publicación.

Si la fuente permite comprobar el contenido general de una declaración, pero no las palabras exactas, deberá utilizarse una paráfrasis y no comillas.

La reconstrucción estilística de una frase mediante IA nunca deberá presentarse como una cita real.

3.7. Verificación de normas, sentencias y documentos jurídicos

La utilización de inteligencia artificial en asuntos jurídicos requiere controles específicos debido a la posibilidad de que el sistema genere disposiciones inexistentes, mezcle versiones normativas o atribuya criterios a decisiones judiciales que no contienen tales conclusiones.

Cuando se utilicen normas, reglamentos, resoluciones, sentencias, autos, expedientes u otros documentos jurídicos, deberán comprobarse directamente:

- la existencia del instrumento;
- su número o identificación;

- la autoridad que lo expidió;
- la fecha;
- su texto pertinente;
- su vigencia o estado procesal, cuando corresponda;
- las reformas o decisiones posteriores que puedan afectarlo; y
- el alcance real de la disposición o criterio citado.

La IA podrá ayudar a localizar, resumir o explicar un documento jurídico, pero la afirmación publicada deberá sustentarse en el documento correspondiente y no únicamente en el resumen generado.

Cuando una conclusión jurídica sea relevante para el contenido, podrá requerirse revisión especializada conforme al nivel de riesgo establecido en este Manual.

3.8. Verificación de cifras y datos cuantitativos

Las cifras deberán tratarse como información que requiere comprobación independiente.

No bastará con confirmar que una cifra aparece en una fuente. Deberá comprenderse, cuando sea relevante, **qué representa esa cifra**.

Esto implica revisar aspectos como:

Universo: a quiénes o qué comprende el dato.

Período: a qué fecha o intervalo corresponde.

Unidad: personas, casos, dólares, porcentajes, eventos u otra medida.

Metodología: cómo fue producido el dato.

Comparabilidad: si puede compararse legítimamente con otras cifras utilizadas.

Actualización: si existe una versión posterior.

Base de cálculo: especialmente cuando la IA haya producido porcentajes, promedios, variaciones o proyecciones.

Cuando la IA realice cálculos, la persona responsable deberá poder reproducir o comprobar razonablemente los resultados relevantes.

Las cifras que sustenten titulares, acusaciones, tendencias o conclusiones de especial importancia deberán recibir revisión reforzada.

3.9. Verificación de imágenes, audio y video

La apariencia visual o auditiva de un contenido no constituye garantía de autenticidad.

Cuando una fotografía, grabación, audio o video sea relevante para demostrar un hecho, deberá intentarse establecer su procedencia y contexto antes de utilizarlo como evidencia.

Dependiendo del riesgo, podrán revisarse elementos como:

origen del archivo, publicación inicial, fecha, ubicación, secuencia completa, versiones anteriores, metadatos disponibles, coincidencias con otros registros y posibles señales de edición o generación sintética.

Las herramientas de inteligencia artificial pueden apoyar este análisis, pero sus conclusiones sobre autenticidad, manipulación o identidad no deberán asumirse como definitivas sin controles adicionales.

Cuando la autenticidad de un material sea central para una publicación de riesgo alto y no pueda ser determinada razonablemente por el equipo, deberá solicitarse revisión técnica especializada.

3.10. Información contradictoria

La existencia de fuentes contradictorias no significa necesariamente que la investigación haya fracasado.

Puede significar que el hecho no se encuentra completamente establecido, que las fuentes tienen acceso a información diferente, que existen interpretaciones legítimamente distintas o que alguna de ellas es incorrecta.

Cuando aparezcan contradicciones relevantes deberá identificarse:

qué sostiene cada fuente; qué evidencia ofrece; cuál es su relación con el hecho; qué interés o posición puede tener; qué aspectos coinciden; y qué elementos permanecen sin resolver.

La IA podrá utilizarse para comparar versiones y señalar discrepancias, pero no deberá decidir automáticamente cuál fuente es verdadera.

Si la evidencia no permite resolver razonablemente la contradicción, la publicación deberá reflejar la incertidumbre en lugar de presentar una conclusión categórica.

3.11. Información que no puede verificarse

No toda información relevante puede ser comprobada mediante documentos públicos.

En periodismo pueden existir testimonios confidenciales, información obtenida directamente de fuentes, denuncias todavía no documentadas u otros elementos cuya verificación requiera métodos distintos.

El hecho de que una información no pueda comprobarse mediante búsqueda digital no la convierte automáticamente en falsa.

Sin embargo, cuando la IA haya sido el único origen del hallazgo y no exista posteriormente ninguna fuente humana, documental o material que permita sostenerlo, la información no deberá utilizarse como hecho.

En estos casos deberá distinguirse entre:

información periodísticamente obtenida por otros medios que la IA no puede verificar, y

afirmaciones generadas o sugeridas por IA para las cuales no existe evidencia adicional.

La primera puede ser evaluada mediante los estándares periodísticos correspondientes. La segunda deberá permanecer fuera del producto factual mientras no exista sustento independiente.

3.12. Verificación negativa: comprobar también lo que no se encontró

En determinadas investigaciones puede ser relevante afirmar que no existe un registro, documento, antecedente o evidencia.

Este tipo de afirmaciones requiere especial cuidado.

Que una herramienta de IA no haya encontrado algo no significa necesariamente que no exista.

Antes de afirmar ausencia deberá evaluarse si:

a) se consultaron las fuentes adecuadas;

b) la búsqueda cubrió el período y territorio correctos;

c) se utilizaron variantes razonables de nombres y términos;

d) la fuente consultada tiene realmente cobertura suficiente; y

e) existen limitaciones técnicas o de acceso que puedan explicar la ausencia.

Cuando la evidencia no permita afirmar inexistencia con suficiente certeza, deberá utilizarse una formulación más precisa, por ejemplo:

“No se encontró registro en las fuentes consultadas”

en lugar de:

“No existe registro.”

3.13. Segunda revisión y principio de independencia

Para información de riesgo alto deberá procurarse que una segunda persona revise los hechos centrales sin limitarse a observar el texto redactado.

Siempre que resulte posible, esa revisión deberá incluir acceso a las fuentes principales.

El revisor deberá poder identificar por sí mismo si la evidencia respalda las afirmaciones realizadas.

Esta revisión adicional busca reducir un riesgo frecuente: que el investigador que realizó la búsqueda se encuentre demasiado familiarizado con su propia hipótesis y pase inadvertidamente por alto errores que una segunda persona pueda detectar.

En productos particularmente sensibles podrá aplicarse una lógica de **doble control**:

investigador verifica → revisor comprueba → responsable editorial aprueba.

No será necesario utilizar este esquema en todos los productos. Su aplicación dependerá del nivel de riesgo y de la importancia de los hechos involucrados.

3.14. Criterios para declarar una información validada

Una información podrá considerarse **validada para uso editorial** cuando, de acuerdo con su naturaleza y nivel de riesgo:

1. exista una fuente o conjunto de fuentes suficientemente identificables;
2. haya sido comprobado que esas fuentes respaldan realmente la afirmación;
3. se haya revisado el contexto necesario para comprender correctamente el dato;
4. se haya determinado que la información se encuentra suficientemente actualizada;
5. se hayan resuelto o identificado las contradicciones relevantes;
6. se hayan aplicado los controles adicionales exigidos para datos personales, información sensible o posibles afectaciones de derechos;
7. se haya realizado la revisión humana correspondiente; y
8. permanezca un nivel de incertidumbre compatible con la forma en que la información será presentada.

Validar no significa alcanzar certeza absoluta.

Significa que, conforme a los estándares profesionales aplicables, existe evidencia suficiente para utilizar la información responsablemente y que las limitaciones conocidas han sido consideradas.

3.15. Validación condicionada

En determinados casos una información podrá utilizarse aun cuando no exista certeza plena, siempre que la incertidumbre sea parte explícita del contenido.

Por ejemplo, podrá señalarse que:

una fuente afirma determinada cuestión; una investigación se encuentra en curso; no fue posible obtener confirmación independiente; existen versiones contrapuestas; determinado dato constituye una estimación; o una conclusión permanece abierta.

En estos casos, la validación no recae necesariamente sobre la verdad definitiva del hecho discutido, sino sobre la exactitud con la cual Fundamedios describe **qué se sabe, quién lo afirma, qué evidencia existe y qué permanece sin comprobar**.

Esta distinción es fundamental para evitar que la inteligencia artificial transforme incertidumbres legítimas en certezas artificiales.

3.16. Registro de la verificación

Cuando el producto tenga riesgo medio o alto, deberá quedar evidencia razonable de cómo se verificaron los elementos centrales.

El registro podrá consistir, según el caso, en enlaces, documentos archivados, capturas, referencias, anotaciones, comentarios editoriales, fichas de verificación, historial institucional de investigación u otro mecanismo que permita reconstruir posteriormente el proceso.

En asuntos de riesgo alto deberá procurarse que pueda identificarse:

qué afirmación se verificó — con qué fuente — quién realizó la comprobación — cuándo — y cuál fue el resultado.

No se exige convertir cada investigación en un expediente administrativo. La documentación deberá ser suficiente para cumplir su finalidad: **permitir que otra persona pueda comprender razonablemente por qué se consideró verificada una afirmación importante.**

3.17. Corrección antes de la producción

Si durante la etapa de verificación se detecta que la IA proporcionó información incorrecta, deberá corregirse el registro de trabajo antes de continuar.

No bastará con que el investigador “recuerde” que determinado dato era falso mientras el resultado incorrecto permanece mezclado con información válida dentro del expediente o conversación de trabajo.

Cuando resulte razonable, deberá marcarse, eliminarse o separar claramente la información descartada para evitar que reaparezca posteriormente durante una nueva etapa de generación.

Esta regla es especialmente importante cuando un mismo entorno de IA conserva contexto de conversaciones anteriores o utiliza documentos acumulados para producir nuevos resultados.

3.18. Cierre de la etapa de verificación

La etapa de verificación concluye cuando los datos necesarios para continuar el trabajo han sido clasificados razonablemente entre:

información validada, que podrá utilizarse para producción;

información condicionada, que podrá utilizarse únicamente expresando adecuadamente sus limitaciones;

información pendiente, que requiere investigación adicional; e

información descartada, que no deberá utilizarse como sustento factual.

La información validada podrá pasar entonces a las etapas de organización, análisis y posterior producción de contenidos.

La regla final será:

Ningún dato material obtenido mediante inteligencia artificial deberá avanzar al proceso de producción editorial si no puede explicarse razonablemente por qué Fundamedios considera que ese dato es suficientemente confiable para el uso que pretende darle.

4. Organización, clasificación y preparación de la información validada

Una vez concluida la etapa de verificación, la información no deberá pasar directamente a la redacción. Antes de ello deberá organizarse de manera que sea posible distinguir con claridad qué hechos han sido comprobados, qué fuentes los respaldan, qué elementos permanecen condicionados o pendientes y qué información no debe utilizarse.

Esta etapa cumple una función práctica y de control. Reduce el riesgo de que, durante la elaboración de un contenido, vuelvan a mezclarse datos confirmados con hipótesis preliminares, referencias descartadas o resultados incorrectos producidos por la inteligencia artificial.

La organización deberá permitir que el equipo pueda trabajar sobre una **base informativa depurada**, especialmente cuando varias personas participen en la investigación o cuando posteriormente se utilice IA para producir borradores.

4.1. Separación entre hechos, fuentes e interpretaciones

La información preparada para la etapa de producción deberá distinguir, en la medida en que corresponda al tipo de trabajo, entre tres elementos:

Hechos verificados, es decir, datos que cuentan con respaldo suficiente y pueden ser utilizados como afirmaciones factuales.

Fuentes, que constituyen el soporte documental, testimonial, audiovisual o de otra naturaleza que permite verificar esos hechos.

Interpretaciones o análisis, que corresponden a conclusiones, lecturas, explicaciones o relaciones construidas a partir de los hechos disponibles.

Esta distinción es especialmente importante cuando la inteligencia artificial ha intervenido en el análisis. El sistema puede identificar patrones, conexiones o explicaciones plausibles que resulten útiles para el investigador, pero una relación sugerida por IA no deberá transformarse automáticamente en un hecho.

Por ejemplo, que dos personas aparezcan mencionadas en varios documentos no permite afirmar por sí solo que exista una relación directa entre ellas. Esa conclusión requerirá evidencia adicional.

4.2. Clasificación mínima de la información

Dependiendo de la complejidad de la investigación, podrá utilizarse una clasificación como la siguiente:

A. Información validada para uso editorial. Puede utilizarse como sustento factual, respetando el contexto y alcance con el que fue verificada.

B. Información condicionada. Puede incorporarse únicamente si se expresa la limitación que la acompaña, por ejemplo, que se trata de una versión no confirmada, una estimación, una denuncia o una interpretación.

C. Información pendiente. No podrá utilizarse como hecho mientras no se completen las verificaciones necesarias.

D. Información descartada. No deberá utilizarse como sustento de la publicación y deberá mantenerse claramente separada para evitar su reutilización accidental.

E. Información sensible o restringida. Podrá encontrarse verificada, pero su utilización o difusión estará sujeta a controles adicionales de protección de datos, confidencialidad, seguridad o revisión jurídica.

Una misma información podrá pertenecer simultáneamente a más de una categoría funcional. Por ejemplo, un dato puede estar validado, pero seguir siendo de acceso restringido por contener información personal sensible.

4.3. Consolidación de las fuentes

Las fuentes relevantes deberán quedar organizadas de modo que puedan ser localizadas nuevamente.

No será suficiente depender exclusivamente del historial de conversación de la herramienta de IA.

Cuando la investigación tenga una complejidad media o alta, deberá procurarse conservar, según corresponda:

- enlace;
- título o identificación del documento;
- institución o autor;
- fecha;
- página, sección, minuto o ubicación relevante;
- copia institucional cuando sea necesario;
- descripción breve de qué información sustenta; y
- cualquier advertencia relevante sobre su uso.

Cuando una fuente pueda desaparecer, modificarse o ser eliminada —por ejemplo, una publicación en redes sociales— deberá evaluarse la conveniencia de conservar una copia, captura o registro institucional suficiente para acreditar su existencia y contexto.

4.4. Preparación de un expediente o repositorio de trabajo

Las investigaciones sustantivas deberán contar con un espacio institucional donde se concentre la información que deba preservarse.

Este repositorio podrá adoptar distintas formas según la naturaleza del proyecto: carpeta institucional, espacio de trabajo, base documental, proyecto dentro del entorno de IA, sistema de gestión de contenidos o una combinación de estos mecanismos.

Lo importante será que la información relevante no quede dispersa únicamente entre:

conversaciones personales, descargas locales, correos individuales, capturas sin identificación o cuentas privadas de herramientas de IA.

El repositorio deberá permitir, en la medida de lo posible, que otra persona autorizada pueda comprender el estado de la investigación y continuarla sin depender exclusivamente de la memoria del investigador original.

4.5. Depuración antes de utilizar IA para producción

Cuando la información recopilada vaya a proporcionarse posteriormente a una herramienta de inteligencia artificial para elaborar un borrador, deberá realizarse una depuración previa.

El conjunto utilizado como input deberá contener únicamente la información pertinente para la tarea.

No deberán incorporarse de forma indiscriminada:

1. datos descartados;

2. hipótesis que ya hayan sido refutadas;

3. información personal innecesaria;

4. documentos completos cuando únicamente se requiera una sección y el resto contenga información sensible;

5. duplicados que puedan distorsionar el peso relativo de una fuente; o

6. información no verificada que pueda confundirse posteriormente con hechos comprobados.

Esta depuración es importante porque la herramienta puede utilizar cualquier contenido proporcionado como parte de su contexto y reproducirlo posteriormente sin reconocer adecuadamente que se trataba de información provisional o descartada.

4.6. Tratamiento de información contradictoria

Cuando existan versiones contradictorias que deban mantenerse dentro del expediente, deberán conservarse de manera que la divergencia resulte visible.

No conviene “resolver” el conflicto eliminando una de las fuentes simplemente porque no coincide con la hipótesis principal de la investigación.

Cuando la contradicción sea relevante para el producto, el material preparado para análisis deberá indicar claramente:

qué sostiene la primera fuente; qué sostiene la segunda; qué evidencia respalda a cada una; y qué aspectos permanecen sin resolver.

La inteligencia artificial podrá utilizarse posteriormente para comparar estas versiones, pero deberá recibir instrucciones explícitas para no presentar como resuelta una contradicción que el expediente mantiene abierta.

4.7. Datos personales e información restringida dentro del expediente

La información deberá organizarse también según su nivel de acceso.

No toda persona que participe en una investigación necesitará conocer todos los datos recopilados.

Cuando existan datos personales sensibles, información confidencial, fuentes protegidas u otros contenidos de acceso restringido, deberán aplicarse mecanismos que permitan separar el material general del material protegido.

Podrán utilizarse, según corresponda:

seudónimos o códigos internos, para evitar identificar innecesariamente a determinadas personas;

carpetas o espacios restringidos, cuando solo una parte del equipo deba tener acceso;

versiones depuradas de documentos, cuando el contenido completo no sea necesario para el trabajo editorial; o

referencias indirectas, cuando resulte suficiente señalar que existe determinada evidencia sin exponer información que no debe circular.

El principio será mantener el mayor nivel de acceso necesario para realizar el trabajo y el menor nivel de exposición posible para la información sensible.

4.8. Registro de procedencia

Cuando se incorporen datos a una base o expediente, deberá procurarse conservar su procedencia.

Esto resulta especialmente importante cuando se combinan múltiples fuentes o cuando la IA extrae información de documentos extensos.

Un dato sin procedencia identificable pierde gran parte de su valor para una investigación que luego debe ser revisada.

Por ello, siempre que sea relevante, deberá poder reconstruirse:

dato → fuente → ubicación dentro de la fuente → proceso de verificación.

Esta relación deberá preservarse incluso cuando la información sea reorganizada en tablas, cronologías, fichas o resúmenes.

4.9. Información derivada por análisis de IA

La organización de información puede producir nuevos resultados que no aparecen literalmente en ninguna fuente.

Por ejemplo, la IA puede:

- identificar que determinados eventos siguen un patrón temporal;
- detectar coincidencias entre contratos;
- agrupar declaraciones por tema;
- calcular tendencias;
- señalar relaciones entre documentos; o
- proponer explicaciones para una secuencia de hechos.

Estos resultados deberán identificarse como **información derivada o analítica**, no como fuente primaria.

Cuando el análisis genere una conclusión relevante para la publicación, deberá comprobarse que pueda reconstruirse a partir de los datos utilizados.

Si una relación identificada por IA no puede reproducirse o explicarse razonablemente, no deberá utilizarse como conclusión sustantiva.

4.10. Control de versiones

Cuando una investigación se desarrolle durante un período prolongado o involucre múltiples revisiones, deberá evitarse trabajar simultáneamente con versiones incompatibles del mismo material.

La organización deberá procurar que exista una versión identificable del expediente o conjunto de datos que se considere vigente para la etapa de producción.

Los cambios sustanciales —por ejemplo, la eliminación de una fuente, corrección de una cifra, incorporación de un nuevo documento o modificación de una conclusión— deberán poder ser conocidos por quienes continúan trabajando sobre el producto.

Esto resulta especialmente importante cuando se utilizan asistentes de IA conectados a repositorios institucionales. Si el sistema continúa accediendo a una versión antigua del material, puede volver a introducir información que ya fue corregida o descartada.

4.11. Preparación para la fase de análisis y producción

Antes de trasladar la información a la siguiente etapa deberá existir un conjunto de trabajo suficientemente ordenado que permita diferenciar:

qué puede afirmarse;

qué debe presentarse con reservas;

qué todavía debe investigarse;

qué no deberá utilizarse;

qué información exige protección reforzada; y

qué fuentes respaldan cada elemento relevante.

No será necesario que todas las investigaciones utilicen exactamente el mismo formato. Un monitoreo diario puede necesitar un sistema mucho más ligero que una investigación extensa de corrupción o una verificación compleja.

El grado de organización deberá ser proporcional al riesgo y volumen de información.

4.12. Condición para avanzar

La información podrá avanzar hacia las fases de análisis y producción cuando:

- los hechos centrales hayan sido verificados;
- las fuentes relevantes puedan localizarse;
- las contradicciones significativas se encuentren identificadas;
- los datos pendientes o descartados estén separados;
- la información sensible cuente con los controles de acceso correspondientes; y
- exista una base suficientemente depurada para evitar que la IA incorpore posteriormente datos que no deberían formar parte del contenido.

A partir de ese momento, la información deja de ser únicamente un conjunto de hallazgos y se convierte en un **insumo institucional preparado para análisis y producción**.

5. Transferencia de la información validada a la fase de producción

La conclusión de la investigación no implica que toda la información recopilada deba trasladarse automáticamente a una herramienta de inteligencia artificial para redactar un contenido.

Antes de iniciar la fase de producción deberá realizarse una **selección consciente del material que será utilizado como insumo**, de manera que el sistema trabaje sobre información depurada, pertinente y suficientemente validada.

La regla general será que la etapa de producción reciba únicamente aquello que el equipo considere necesario para elaborar el producto correspondiente.

5.1. Preparación del paquete de información

Cuando se utilice IA para redactar o estructurar un contenido, el material proporcionado deberá organizarse de forma que permita distinguir con claridad:

Hechos confirmados. Información que puede incorporarse como afirmación factual.

Información condicionada. Datos que solo pueden utilizarse si se conserva expresamente la advertencia o limitación correspondiente.

Contexto. Antecedentes que permiten comprender adecuadamente el asunto, pero que no necesariamente constituyen el núcleo del producto.

Citas verificadas. Declaraciones cuya literalidad y atribución han sido comprobadas.

Fuentes de respaldo. Documentos, enlaces, registros o referencias que permiten revisar los hechos incluidos.

Restricciones editoriales. Elementos que no deben afirmarse, inferirse o presentarse como concluyentes.

La IA deberá recibir, en la medida de lo posible, instrucciones que reflejen estas diferencias.

No deberá proporcionarse un conjunto indiferenciado de documentos esperando que el sistema determine por sí solo qué información es confiable, qué fue descartado durante la investigación o qué afirmaciones pueden publicarse.

5.2. Exclusión de información descartada o innecesaria

Antes de trasladar información a la fase de producción deberán excluirse los elementos que hayan sido identificados como falsos, incorrectos, irrelevantes o insuficientemente sustentados.

También deberá eliminarse del input aquella información que, aunque sea correcta, no resulte necesaria para el producto y cuya incorporación genere riesgos adicionales de privacidad, confidencialidad o distorsión.

Esta exclusión es especialmente importante cuando se utilizan sistemas capaces de conservar contexto o trabajar sobre grandes repositorios. Un dato incorrecto que permanezca disponible para la herramienta puede reaparecer posteriormente en una redacción distinta y parecer nuevamente válido.

Por ello, **la depuración de la información forma parte del control editorial y no únicamente de la organización documental**.

5.3. Conservación de incertidumbres

Cuando una investigación haya concluido que determinado aspecto permanece abierto, controvertido o no suficientemente comprobado, esta circunstancia deberá trasladarse expresamente a la fase de producción.

No deberá pedirse a la IA que “complete” una conclusión cuando el expediente no permita hacerlo.

Si existen dos versiones relevantes y ninguna ha sido definitivamente acreditada, el sistema deberá conocer esa limitación y recibir instrucciones para mantenerla.

De igual forma, cuando una cifra sea una estimación, una declaración corresponda únicamente a una de las partes o determinado elemento no haya podido confirmarse de forma independiente, dicha condición deberá acompañar al dato durante el proceso de redacción.

5.4. Correspondencia entre insumos y producto final

La persona responsable deberá poder comprobar que las afirmaciones incluidas posteriormente en el producto se encuentran dentro del universo de información validada que fue autorizado para producción.

La herramienta de IA no deberá introducir como hecho:

a) información proveniente de su conocimiento general que no haya sido requerida y verificada;

b) antecedentes nuevos que no formen parte del expediente;

c) citas no proporcionadas por el equipo;

d) cifras obtenidas por inferencia;

e) nombres, cargos, fechas o relaciones añadidos para “completar” el texto; ni

f) conclusiones que no se desprendan razonablemente de los insumos autorizados.

Cuando durante la redacción la IA introduzca un dato nuevo que pudiera resultar relevante, este deberá regresar a la fase de investigación y verificación antes de poder incorporarse definitivamente.

En otras palabras:

la producción no reemplaza la investigación. Si aparece información nueva durante la redacción, esa información vuelve al circuito de verificación.

5.5. Separación entre expediente completo e input de producción

El expediente de investigación podrá contener más información que la necesaria para redactar el producto final.

No deberá asumirse que todo el expediente debe cargarse automáticamente en la herramienta de IA.

En particular, cuando existan fuentes confidenciales, datos sensibles, documentos reservados o información cuya revelación no sea necesaria para la redacción, deberá prepararse un **input de producción depurado** que contenga únicamente los elementos indispensables.

Este mecanismo permite conservar un expediente completo para fines de verificación y, al mismo tiempo, reducir la exposición innecesaria de información durante la generación de contenidos.

5.6. Transferencia dentro del entorno institucional

La información validada deberá pasar a la fase de producción utilizando los entornos, repositorios y herramientas institucionales definidos por Fundamedios.

Cuando la producción sea realizada por otra persona, deberá entregarse el material de manera que pueda comprender:

qué está comprobado, qué permanece condicionado, cuáles son las fuentes principales y qué restricciones deben respetarse.

La transferencia no deberá depender exclusivamente de una explicación verbal o de la continuidad de una conversación personal con una herramienta de IA.

En investigaciones relevantes, la información necesaria deberá quedar disponible dentro del repositorio institucional correspondiente para permitir continuidad, revisión y trazabilidad.

5.7. Punto de control antes de la producción

Antes de iniciar la generación de un borrador sustantivo mediante inteligencia artificial, la persona responsable deberá comprobar que:

- el objeto del producto se encuentra definido;
- los hechos centrales han sido validados;
- las fuentes principales se encuentran identificadas;
- las contradicciones relevantes están documentadas;
- los elementos pendientes se encuentran claramente diferenciados;
- la información descartada no forma parte de los insumos;
- los datos personales y contenidos confidenciales han sido reducidos al mínimo necesario;
- las citas que vayan a utilizarse se encuentran verificadas; y
- la herramienta seleccionada está autorizada para procesar la información proporcionada.

Cuando alguno de estos elementos sea indispensable y todavía no se encuentre resuelto, la producción deberá posponerse o limitarse a tareas que no dependan de esa información.

5.8. Cierre del Capítulo II

El proceso de investigación asistida por inteligencia artificial concluye cuando Fundamedios cuenta con un conjunto de información **identificada, verificada, clasificada, organizada y preparada para su utilización posterior**.

A partir de ese momento, la inteligencia artificial deja de funcionar principalmente como herramienta de búsqueda y pasa a una función distinta: **apoyar el análisis, estructuración y producción de contenidos a partir de información previamente controlada**.

La separación entre ambas fases constituye uno de los controles fundamentales de este Manual:

primero se determina qué información puede utilizarse; después se utiliza la inteligencia artificial para trabajar con ella.

La generación de un texto nunca deberá ser el mecanismo mediante el cual se decide, por primera vez, si la información que lo sustenta era correcta.

CAPÍTULO

III

Gestión, almacenamiento y protección de la información



La utilización de inteligencia artificial modifica la forma en que la información circula dentro de una organización. Un documento puede ser cargado en una herramienta, procesado, resumido, fragmentado, combinado con otras fuentes y posteriormente incorporado a nuevas conversaciones, bases documentales o productos. Cada una de estas operaciones puede generar nuevas copias, registros o formas derivadas de la información original.

Por ello, la protección no deberá concentrarse únicamente en el documento inicial. Deberá comprender **todo el ciclo de vida de la información utilizada o generada mediante inteligencia artificial**.

Fundamedios administrará este ciclo procurando que la información permanezca bajo control institucional, que únicamente sea accesible por quienes razonablemente la necesitan, que no se conserve más allá de lo necesario y que su tratamiento respete las obligaciones aplicables en materia de protección de datos, confidencialidad y seguridad.

1. Gestión institucional del ciclo de información

La información utilizada en procesos asistidos por inteligencia artificial deberá gestionarse desde su incorporación al entorno institucional hasta su conservación o eliminación.

Para efectos de este Manual, el ciclo comprende:

ingreso → clasificación → utilización → transformación → almacenamiento → acceso → reutilización → archivo o eliminación

La existencia de IA dentro de este ciclo no modifica la naturaleza de la información. Un dato personal continúa siendo un dato personal cuando se incorpora a un prompt; un documento confidencial no pierde esa condición porque haya sido resumido; y una fuente protegida puede seguir siendo identificable aun cuando su nombre haya sido eliminado si permanecen suficientes elementos para reconocerla.

1.1. La información institucional debe permanecer bajo control institucional

La información obtenida o producida en actividades desarrolladas para Fundamedios deberá permanecer, cuando corresponda conservarla, dentro de los entornos y repositorios institucionales definidos para ese propósito.

Esta regla comprende tanto los documentos originales como los materiales que puedan generarse durante el uso de IA: archivos de trabajo, bases estructuradas, transcripciones, resúmenes, clasificaciones, prompts que formen parte de una metodología institucional, resultados relevantes, verificaciones y demás elementos necesarios para reconstruir o continuar un proceso.

El objetivo no será almacenar indiscriminadamente todas las interacciones realizadas con inteligencia artificial. Se conservará aquello que tenga valor institucional, documental, metodológico, probatorio o de trazabilidad conforme a las reglas establecidas en este Manual.

La información que deba conservarse **no deberá depender exclusivamente de la cuenta, dispositivo o espacio personal de quien realizó el trabajo**.

1.2. Clasificación según el nivel de protección

Antes de decidir dónde almacenar información o mediante qué herramienta podrá procesarse, deberá conocerse razonablemente su nivel de protección.

Fundamedios podrá organizar la información institucional bajo las siguientes categorías generales:

Pública. Información destinada al conocimiento público o cuya divulgación no genera restricciones relevantes para la organización.

Interna. Información de trabajo que no se encuentra destinada a difusión pública, pero cuyo acceso dentro de la organización no representa normalmente un riesgo significativo.

Confidencial. Información cuyo acceso deberá limitarse debido a su contenido, a obligaciones asumidas por Fundamedios o a los posibles efectos de su divulgación.

Restringida. Información cuya exposición indebida pueda generar un riesgo significativo para personas, fuentes, investigaciones, sistemas o para la propia organización y que, por ello, requiere los controles más elevados de acceso y procesamiento.

La clasificación deberá responder al contenido y no únicamente al nombre o ubicación del archivo.

Un documento aparentemente ordinario puede contener dentro de sus anexos información que exija una protección superior. Del mismo modo, una base elaborada a partir de fuentes públicas puede requerir mayores controles cuando la agregación de esos datos permita obtener información nueva sobre personas identificables.

1.3. Qué información deberá considerarse especialmente protegida

Sin perjuicio de la clasificación que corresponda en cada caso, requerirán particular atención:

- datos personales sensibles y otras categorías de datos sujetas a protección reforzada;
- identidad de fuentes confidenciales y elementos que permitan inferirla;
- información sobre víctimas, denunciantes, menores de edad o personas cuya situación pueda generar riesgos especiales;
- investigaciones todavía no publicadas;
- información recibida bajo compromisos de confidencialidad;
- estrategias jurídicas y comunicaciones protegidas;
- credenciales, información de autenticación y datos relacionados con la seguridad de los sistemas;
- documentos cuyo acceso se encuentre legal o contractualmente restringido; y
- información cuya divulgación anticipada pueda comprometer una investigación o generar un riesgo relevante para una persona.

La clasificación deberá considerar no solamente el posible perjuicio para Fundamedios, sino también las consecuencias que una exposición pueda tener para terceros.

1.4. La transformación mediante IA no elimina la protección

Cuando una herramienta de inteligencia artificial resuma, traduzca, clasifique, extraiga o reorganice información protegida, el resultado deberá conservar un nivel de protección compatible con el contenido que todavía permita revelar.

Por ejemplo, un resumen de una entrevista con una fuente confidencial puede continuar siendo restringido aunque el archivo de audio original no se encuentre adjunto. Una cronología construida a partir de expedientes con datos personales puede continuar constituyendo tratamiento de esos datos. Una traducción de un documento reservado conserva la misma naturaleza del documento del cual proviene.

La protección deberá evaluarse sobre el **contenido resultante**, no únicamente sobre su formato.

1.5. Información agregada e inferencias

La inteligencia artificial permite relacionar información que, examinada aisladamente, podría parecer inocua.

Por esta razón deberá considerarse también el riesgo derivado de la **agregación**.

La combinación de diferentes datos públicos puede permitir identificar patrones de comportamiento, relaciones personales, ubicaciones, rutinas, condiciones económicas u otros aspectos que no resultaban evidentes en cada fuente individual.

De igual manera, una herramienta puede producir inferencias sobre una persona a partir de los datos proporcionados.

Estas inferencias no deberán considerarse hechos simplemente porque hayan sido obtenidas mediante análisis automatizado y, cuando se refieran a personas identificadas o identificables, deberán someterse a las reglas de protección y verificación correspondientes.

1.6. Acceso según necesidad

El hecho de pertenecer a Fundamedios o participar en un proyecto no implicará acceso irrestricto a toda la información disponible dentro de sus sistemas de inteligencia artificial.

Los accesos deberán asignarse de acuerdo con las funciones y necesidades de cada persona.

En términos generales:

quien necesita consultar, podrá consultar;

quien necesita modificar, podrá modificar;

quien necesita administrar, tendrá permisos de administración;

quien no necesita acceder, no deberá recibir acceso.

Esta separación deberá aplicarse especialmente a proyectos sensibles, investigaciones no publicadas, datos personales, fuentes protegidas y repositorios que concentren grandes cantidades de información.

1.7. Copias y duplicación de información

La facilidad para copiar información entre herramientas puede provocar una proliferación innecesaria de archivos y versiones.

Deberá evitarse descargar, copiar, reenviar o volver a cargar información cuando no exista una necesidad concreta.

Cuando un documento pueda consultarse desde un repositorio institucional autorizado, deberá preferirse mantener una referencia controlada antes que generar múltiples copias en dispositivos, cuentas o plataformas distintas.

En especial, no deberán crearse repositorios paralelos de información institucional dentro de cuentas personales de inteligencia artificial, almacenamiento personal en la nube u otros servicios no autorizados.

1.8. Uso posterior de la información

La información recopilada para una investigación determinada no deberá considerarse automáticamente disponible para cualquier utilización futura.

Antes de reutilizar información —especialmente cuando contenga datos personales, información confidencial o material obtenido bajo condiciones particulares— deberá considerarse si la nueva finalidad es compatible con aquella que justificó inicialmente su obtención.

Esto adquiere especial importancia cuando se construyan repositorios institucionales destinados a ser consultados posteriormente mediante inteligencia artificial.

El hecho de que Fundamedios posea legítimamente determinada información **no significa necesariamente que todo su contenido deba incorporarse al conocimiento permanente de una herramienta de IA**.

La selección de información destinada a repositorios de consulta, memoria institucional, bases de conocimiento o sistemas de recuperación deberá realizarse de forma deliberada.

1.9. Memoria institucional de inteligencia artificial

Fundamedios podrá desarrollar progresivamente una base de conocimiento institucional destinada a mejorar la capacidad de sus herramientas de inteligencia artificial para asistir el trabajo de la organización.

Esta memoria podrá incluir, según resulte apropiado, manuales, metodologías, criterios editoriales, procedimientos, documentos de referencia, investigaciones publicadas, formatos, terminología institucional, antecedentes de proyectos, fuentes de consulta habitual y otros materiales que tengan utilidad transversal.

Su construcción deberá realizarse bajo un criterio de **selección**, y no mediante la acumulación automática de todo aquello que haya pasado por una herramienta.

Antes de incorporar información a una memoria o base de conocimiento permanente deberá considerarse:

a) si existe una finalidad institucional razonable para conservarla;

b) si la información se encuentra suficientemente validada;

c) si contiene datos que deban eliminarse, anonimizarse o restringirse;

d) quién podrá acceder posteriormente;

e) durante cuánto tiempo resulta razonable conservarla; y

f) qué consecuencias tendría que la herramienta la utilizara en contextos futuros distintos de aquel en que fue originalmente producida.

La información falsa, descartada, desactualizada o cuya confiabilidad no haya podido establecerse no deberá incorporarse deliberadamente como conocimiento institucional validado.

1.10. Diferencia entre archivo y memoria de IA

Deberá distinguirse entre **conservar información** y **hacerla disponible permanentemente a una inteligencia artificial**.

Un expediente puede necesitar conservarse durante años por razones periodísticas, jurídicas, institucionales o documentales sin que exista ninguna necesidad de incorporarlo a la memoria activa de un sistema.

El archivo institucional tiene como finalidad preservar información.

La memoria o base de conocimiento de IA tiene una finalidad distinta: **permitir que determinada información pueda ser recuperada y utilizada por el sistema en trabajos posteriores**.

Por ello, el acceso de la IA a un repositorio deberá ser una decisión específica y controlada. No todos los archivos institucionales deberán formar parte automáticamente de su base de conocimiento.

1.11. Corrección del conocimiento institucional

Cuando se determine que una información incorporada a una base de conocimiento institucional es incorrecta, se encuentra desactualizada o ha sido sustituida por una versión posterior, deberá procurarse su corrección.

No será suficiente corregir únicamente el producto final si la herramienta continúa teniendo acceso a la información errónea que originó el problema.

Según las capacidades del sistema, podrá ser necesario sustituir un documento, actualizar una fuente, modificar una entrada, retirar información del repositorio o establecer una advertencia que permita reconocer que determinada versión ya no debe utilizarse.

La administración del conocimiento institucional deberá procurar que la IA no reproduzca en nuevos trabajos errores que la organización ya haya identificado y corregido.

1.12. Responsabilidad sobre el ciclo de información

Toda persona que utilice información institucional mediante IA será responsable de aplicar los controles correspondientes durante la parte del ciclo en la que intervenga.

La administración general de los repositorios, permisos, herramientas y políticas corresponderá a las personas o unidades que Fundamedios designe para ello.

Cuando una persona detecte que información se encuentra almacenada en un lugar inadecuado, que existen permisos excesivos, que un documento sensible fue incorporado indebidamente a una herramienta o que una base de conocimiento contiene información que no debería estar disponible, deberá comunicarlo para que se adopten las medidas correspondientes.

La gestión adecuada de la información no termina cuando se publica una nota. Continúa mientras Fundamedios conserve los datos, documentos, registros o conocimiento generado durante el proceso.

2. Protección de datos personales en el uso de inteligencia artificial

El uso de inteligencia artificial puede implicar tratamiento de datos personales incluso cuando esa no sea la finalidad principal de una investigación. Nombres, fotografías, voces, direcciones, identificadores digitales, antecedentes profesionales, publicaciones en redes sociales, documentos judiciales, bases de datos, entrevistas y otros materiales utilizados habitualmente en actividades periodísticas pueden contener información relacionada con personas identificadas o identificables.

La protección deberá observarse durante todo el proceso. No comienza al momento de publicar ni se limita a evitar la difusión de determinados datos. Comprende su **obtención, ingreso a una herramienta de IA, análisis, combinación, almacenamiento, utilización, transferencia, conservación y eliminación**.

Fundamedios aplicará en sus actividades la normativa ecuatoriana vigente sobre protección de datos personales y, particularmente, los principios y obligaciones pertinentes de la Ley Orgánica de Protección de Datos Personales, su Reglamento General y las disposiciones emitidas por la autoridad competente. Su aplicación deberá realizarse conjuntamente con las garantías constitucionales relacionadas con privacidad, intimidad, protección de datos y libertad de expresión, así como con las reglas especiales que correspondan al ejercicio de actividades periodísticas.

El tratamiento de datos personales mediante inteligencia artificial deberá partir de una regla práctica: **que una herramienta permita técnicamente procesar determinada información no significa que Fundamedios deba proporcionársela ni que pueda utilizarla para cualquier finalidad.**

2.1. Identificación de datos personales

Antes de proporcionar información a una herramienta de inteligencia artificial deberá considerarse si el material contiene datos personales.

La identificación no deberá limitarse a nombres o números de identificación. Un dato puede permitir identificar a una persona indirectamente cuando se combina con otros elementos.

Entre la información que puede constituir datos personales se encuentran, según el contexto:

- nombres, apellidos, números de identificación y datos de contacto;
- fotografías, grabaciones de voz y videos;
- ubicación y datos asociados a desplazamientos;
- direcciones IP, identificadores digitales y determinados metadatos;
- información laboral, profesional, patrimonial o económica;
- publicaciones, comentarios y actividades atribuibles a una persona;
- relaciones familiares o personales;
- información contenida en expedientes, denuncias, contratos y registros;
- perfiles, clasificaciones e inferencias producidas sobre una persona; y
- combinaciones de información que permitan razonablemente identificarla aunque su nombre haya sido eliminado.

La evaluación deberá atender al contenido real y al contexto. Sustituir un nombre por expresiones como “el funcionario”, “la denunciante” o un código interno no implica necesariamente que la información haya dejado de ser personal si la persona continúa siendo identificable mediante los demás datos disponibles.

2.2. Finalidad del tratamiento

Antes de procesar datos personales mediante IA deberá existir una finalidad determinada y vinculada con las actividades de Fundamedios.

La información recopilada para investigar un hecho no deberá utilizarse automáticamente para elaborar perfiles sobre las personas involucradas, alimentar bases de conocimiento permanentes o realizar análisis diferentes que no guarden una relación razonable con el propósito original.

Cuando una herramienta permita efectuar nuevos análisis sobre información ya disponible, deberá preguntarse primero:

¿Para qué necesitamos este nuevo tratamiento y qué relación tiene con el trabajo que estamos realizando?

La facilidad tecnológica para reutilizar datos no constituye por sí sola una finalidad legítima.

Si la finalidad cambia sustancialmente, deberá evaluarse nuevamente la necesidad, proporcionalidad y condiciones bajo las cuales puede realizarse el tratamiento.

2.3. Minimización de datos

Fundamedios deberá utilizar únicamente los datos personales que resulten adecuados, pertinentes y necesarios para la finalidad perseguida.

Este principio adquiere especial importancia al trabajar con inteligencia artificial, porque existe una tendencia práctica a proporcionar al sistema documentos completos o grandes conjuntos de información aunque la tarea requiera únicamente una parte.

Antes de cargar un documento, base de datos, transcripción o expediente deberá considerarse si puede realizarse el mismo trabajo proporcionando menos información.

Por ejemplo, si la IA debe comparar la estructura de dos contratos, puede no necesitar nombres, números de identificación, firmas o datos bancarios de las partes. Si debe clasificar testimonios por tema, puede no necesitar conocer la identidad de quienes los proporcionaron. Si debe resumir determinados argumentos de un expediente, puede resultar innecesario cargar anexos que contienen información personal ajena al análisis.

La regla será **proporcionar a la herramienta lo necesario para realizar la tarea y excluir aquello que no necesita conocer**.

2.4. Anonimización y seudonimización

Cuando la finalidad pueda alcanzarse sin identificar directamente a las personas, deberán utilizarse mecanismos que reduzcan la exposición de sus datos.

La **seudonimización** podrá consistir, por ejemplo, en sustituir nombres por códigos o referencias internas. La información continúa siendo personal cuando Fundamedios conserva la posibilidad de relacionar posteriormente el código con la persona.

La **anonimización**, en cambio, deberá impedir razonablemente que la información pueda volver a asociarse con una persona identificada o identificable.

Esta diferencia deberá tenerse presente porque la mera eliminación del nombre no convierte automáticamente un conjunto de datos en anónimo.

Cuando se utilicen códigos, la información que permita restablecer la identidad deberá conservarse separadamente y con acceso limitado cuando la naturaleza del trabajo así lo requiera.

2.5. Datos sensibles y categorías que requieren protección reforzada

Determinada información puede generar consecuencias particularmente graves si es utilizada, inferida o divulgada indebidamente. Su tratamiento mediante inteligencia artificial deberá recibir una protección reforzada y sujetarse estrictamente a la normativa aplicable.

Cuando una investigación involucre datos sensibles o información equivalente sometida a protección especial, deberá evaluarse previamente si su procesamiento mediante IA es realmente necesario.

La misma precaución se aplicará cuando la IA pueda **inferir** información sensible a partir de datos que, individualmente, no parecen tener esa naturaleza.

Un sistema puede, por ejemplo, sugerir características o condiciones personales basándose en patrones de comportamiento, lenguaje, ubicación, relaciones o publicaciones. Estas inferencias deberán tratarse con especial cautela y no deberán presentarse como hechos sin evidencia suficiente.

Cuando la información sensible no resulte indispensable para la tarea, deberá excluirse del input.

2.6. Información pública y datos personales

La circunstancia de que un dato pueda encontrarse públicamente en Internet no elimina automáticamente las obligaciones relacionadas con su tratamiento.

Existe una diferencia entre que determinada información pueda ser consultada individualmente en una fuente pública y que sea recopilada masivamente, combinada con otras bases, sometida a análisis automatizado o utilizada para construir perfiles o inferencias sobre una persona.

Por ello, al utilizar IA para investigar información pública deberá considerarse no solamente **de dónde proviene el dato**, sino también **qué pretende hacer Fundamedios con él**.

La naturaleza pública de la fuente será un elemento relevante para la evaluación, pero no deberá utilizarse como autorización general para cualquier procesamiento posterior.

2.7. Datos personales y actividad periodística

La protección de datos personales deberá aplicarse de manera compatible con el ejercicio de las libertades de expresión, información y prensa y con las reglas aplicables a la actividad periodística.

Este equilibrio deberá analizarse especialmente cuando la información tenga relevancia pública o resulte necesaria para investigar asuntos de interés general.

La existencia de datos personales dentro de una investigación no significa, por sí sola, que estos no puedan ser tratados o publicados. Del mismo modo, la existencia de interés periodístico no elimina automáticamente toda consideración sobre privacidad.

La evaluación deberá atender a elementos como **la relevancia pública de la información, la posición o actuación pública de la persona, la necesidad del dato para comprender el hecho investigado, el posible daño derivado de su tratamiento, la existencia de personas vulnerables y la posibilidad de alcanzar la finalidad periodística exponiendo menos información personal**.

En asuntos complejos o de riesgo alto deberá solicitarse revisión jurídica cuando exista una duda razonable sobre el equilibrio entre protección de datos y ejercicio de las libertades informativas.

2.8. Tratamiento de datos de personas vulnerables

Deberá aplicarse especial cautela cuando la información corresponda a niños, niñas y adolescentes, víctimas, denunciantes, testigos, personas sometidas a amenazas u otras personas respecto de las cuales la exposición de información pueda incrementar un riesgo existente.

En estos casos, el análisis no deberá limitarse a determinar si el dato es verdadero.

También deberá preguntarse si es necesario que la herramienta de IA conozca esa información, quién podrá acceder posteriormente al resultado y qué consecuencias podría producir una exposición accidental.

Cuando sea posible, deberán utilizarse versiones depuradas, identificadores internos u otros mecanismos que reduzcan la exposición.

2.9. Inferencias, perfiles y clasificación de personas

La inteligencia artificial puede utilizar información existente para clasificar personas, detectar patrones o generar inferencias sobre su comportamiento, relaciones o características.

Estas capacidades requieren un control particular.

Una inferencia algorítmica no deberá considerarse equivalente a un hecho comprobado.

Cuando la IA sugiera, por ejemplo, que determinadas personas se encuentran relacionadas, que presentan un patrón común de comportamiento o que pertenecen a una categoría determinada, deberá identificarse claramente que se trata de un **resultado analítico que requiere comprobación**.

Fundamedios no deberá utilizar inteligencia artificial para elaborar perfiles invasivos sobre personas cuando estos no resulten necesarios y proporcionales para una finalidad periodística o institucional legítima.

La clasificación automatizada tampoco deberá utilizarse para atribuir ilícitos, intenciones, culpabilidad o características personales sin evidencia independiente suficiente.

2.10. Decisiones automatizadas

La inteligencia artificial no deberá adoptar autónomamente decisiones que produzcan consecuencias significativas para personas dentro de las actividades de Fundamedios.

En particular, un sistema no deberá decidir por sí solo que una persona:

a) ha realizado una afirmación falsa;

b) debe ser objeto de una investigación;

c) se encuentra vinculada con una actividad ilícita;

d) representa determinado nivel de riesgo;

e) debe ser excluida de una actividad o tratamiento institucional; o

f) debe ser objeto de una conclusión editorial adversa.

La IA podrá identificar elementos que merezcan revisión, ordenar información o sugerir posibles patrones, pero la decisión correspondiente deberá ser adoptada mediante intervención humana y sobre evidencia suficiente.

Mientras mayor sea el posible impacto sobre una persona, mayor deberá ser la independencia y profundidad de esa revisión.

2.11. Proveedores de inteligencia artificial y tratamiento por terceros

Cuando Fundamedios proporcione datos personales a una plataforma externa de inteligencia artificial deberá considerar que la información puede estar siendo procesada mediante infraestructura y servicios de terceros.

Antes de autorizar una herramienta para tratamientos relevantes deberá revisarse, de acuerdo con el riesgo, qué ocurre con la información proporcionada: quién la procesa, para qué finalidades, durante cuánto tiempo se conserva, si puede utilizarse para entrenamiento o mejora de modelos, qué terceros pueden intervenir y qué mecanismos existen para su eliminación o control.

Las configuraciones institucionales deberán procurar impedir, cuando las características del servicio lo permitan, que la información proporcionada por Fundamedios sea utilizada para entrenar modelos de uso general o para finalidades ajenas a la prestación del servicio contratado.

Cuando las condiciones ofrecidas por una herramienta sean incompatibles con el nivel de protección requerido, **la información no deberá proporcionarse a esa herramienta**, independientemente de su utilidad funcional.

2.12. Transferencias y procesamiento internacional

Las herramientas de inteligencia artificial pueden procesar o almacenar información en diferentes jurisdicciones.

Cuando el tratamiento implique transferencias o comunicaciones internacionales de datos personales, deberán observarse las condiciones establecidas por la normativa ecuatoriana aplicable.

La evaluación deberá considerar, cuando corresponda, la ubicación del proveedor y de sus subencargados, las condiciones contractuales aplicables, las garantías ofrecidas para el tratamiento y los mecanismos jurídicos utilizados para las transferencias.

En tratamientos de riesgo alto o cuando se procesen volúmenes relevantes de datos personales, deberá existir una evaluación previa suficiente antes de habilitar la herramienta.

2.13. Memoria de IA y datos personales

La posibilidad de que una herramienta recuerde información entre conversaciones puede resultar útil para construir conocimiento institucional, pero introduce un riesgo particular respecto de los datos personales.

Los datos de una investigación no deberán incorporarse automáticamente a una memoria institucional permanente.

Antes de hacerlo deberá determinarse:

1. si existe una necesidad real de conservar esos datos para usos futuros;
2. si la finalidad de la conservación resulta compatible con su obtención;
3. si pueden conservarse de forma anonimizada o menos identificable;
4. quién tendrá acceso a la memoria;
5. durante cuánto tiempo deberán permanecer disponibles; y
6. cómo podrán corregirse o eliminarse posteriormente.

Una base de conocimiento institucional deberá procurar almacenar **conocimiento útil para la organización**, no acumular indiscriminadamente información personal sobre las personas que han aparecido en sus investigaciones.

2.14. Exactitud y actualización de los datos personales

Cuando Fundamedios conserve datos personales que puedan utilizarse posteriormente mediante IA deberá adoptar medidas razonables para evitar que información incorrecta o desactualizada continúe generando nuevos resultados.

Esto resulta especialmente importante cuando una investigación haya sido corregida, una acusación haya sido desmentida, una situación jurídica haya cambiado o un dato inicialmente válido haya dejado de estar vigente.

Si se corrige únicamente la publicación, pero el repositorio conectado a la IA mantiene la información anterior sin ninguna advertencia, el sistema puede volver a utilizarla en el futuro.

Por ello, las correcciones relevantes deberán alcanzar, cuando corresponda, tanto al producto público como a los repositorios internos que puedan alimentar nuevos procesos.

2.15. Conservación y eliminación

Los datos personales no deberán conservarse indefinidamente únicamente porque su almacenamiento sea técnicamente sencillo.

Los períodos de conservación deberán responder a la finalidad, naturaleza de la información, necesidades periodísticas o documentales, obligaciones jurídicas y riesgos asociados.

Al concluir la finalidad correspondiente deberá evaluarse si la información:

debe conservarse, por existir una razón legítima para ello;

debe archivarse con acceso restringido, porque continúa teniendo valor pero no requiere disponibilidad ordinaria;

debe anonimizarse, cuando pueda conservarse su utilidad sin mantener la identificación; o

debe eliminarse, cuando ya no exista una finalidad suficiente para conservarla.

Las reglas específicas de retención y eliminación serán desarrolladas en el apartado correspondiente de este capítulo.

2.16. Derechos de los titulares y capacidad de localización

La utilización de IA no deberá impedir que Fundamedios pueda atender, cuando jurídicamente corresponda, los derechos reconocidos a los titulares de datos personales.

Para ello resulta importante conocer razonablemente dónde se encuentra almacenada la información y en qué sistemas puede haber sido incorporada.

La dispersión de datos entre cuentas personales, múltiples plataformas y conversaciones no administradas dificulta tanto la seguridad como el ejercicio de derechos.

El entorno institucional deberá permitir, dentro de las posibilidades técnicas razonables, localizar, corregir, actualizar, restringir o eliminar información cuando exista obligación de hacerlo.

Cuando una solicitud relacionada con datos personales afecte material utilizado en una investigación o publicación periodística, deberá analizarse también el régimen jurídico aplicable a esa actividad antes de modificar o eliminar información.

2.17. Incidentes relacionados con datos personales

Cualquier utilización de IA que pueda haber producido pérdida, acceso no autorizado, exposición, alteración, comunicación indebida o tratamiento irregular de datos personales deberá tratarse como un posible incidente.

La persona que lo detecte no deberá intentar ocultarlo, resolverlo únicamente eliminando la conversación o esperar a comprobar por sí misma que efectivamente se produjo un daño.

Deberá comunicarlo inmediatamente mediante el procedimiento interno correspondiente para permitir:

- contener el incidente;
- identificar qué información estuvo involucrada;
- determinar qué personas pueden haber sido afectadas;
- conocer qué herramienta o proveedor intervino;
- preservar la evidencia necesaria;
- adoptar medidas de corrección; y
- determinar si existen obligaciones adicionales de comunicación, notificación o actuación conforme a la normativa aplicable.

La rapidez de la comunicación interna será especialmente importante cuando exista posibilidad de limitar el acceso, eliminar información de una plataforma o revocar credenciales.

2.18. Evaluación reforzada para tratamientos de alto riesgo

Cuando un proyecto de inteligencia artificial implique un tratamiento de datos personales que, por su naturaleza, volumen, finalidad, tecnología utilizada o posibles consecuencias, pueda generar un riesgo elevado para los derechos de las personas, no deberá iniciarse como una actividad ordinaria.

Deberá realizarse previamente una evaluación específica que permita comprender el tratamiento proyectado y determinar si los riesgos pueden ser razonablemente controlados.

Esta evaluación deberá considerar, al menos:

Naturaleza del tratamiento: qué información será utilizada y de dónde proviene.

Finalidad: qué pretende obtener Fundamedios y por qué requiere IA.

Necesidad y proporcionalidad: si existe una alternativa menos invasiva.

Funcionamiento: qué hará la herramienta con los datos.

Proveedores: quiénes tendrán acceso o participarán en el procesamiento.

Riesgos: qué podría ocurrir si la información es incorrecta, expuesta, reutilizada o interpretada indebidamente.

Controles: qué medidas técnicas, organizativas y humanas reducirán esos riesgos.

Conservación: qué información permanecerá después de concluir el proyecto.

Cuando el riesgo residual continúe siendo incompatible con los estándares establecidos en este Manual o con la normativa aplicable, el tratamiento no deberá realizarse en la forma propuesta.

2.19. Protección de datos desde el diseño

Los proyectos que incorporen inteligencia artificial deberán procurar resolver los riesgos de protección de datos **antes de comenzar el tratamiento**, y no únicamente cuando aparezca un problema.

Esto implica diseñar desde el inicio cuestiones como qué datos serán necesarios, quién podrá acceder, qué herramienta podrá procesarlos, qué información deberá anonimizarse, dónde quedarán almacenados los resultados, cuánto tiempo se conservarán y cómo podrán eliminarse posteriormente.

En términos operativos, antes de iniciar un uso relevante de IA que involucre datos personales deberá poder responderse:

¿Qué datos necesitamos? ¿Por qué los necesitamos? ¿Puede hacerse con menos? ¿Quién tendrá acceso? ¿Dónde se procesarán? ¿Qué quedará almacenado? ¿Durante cuánto tiempo? ¿Qué ocurriría si esa información se expone o se utiliza incorrectamente?

Si estas preguntas revelan un riesgo que no puede controlarse razonablemente, deberá modificarse el procedimiento antes de continuar.

2.20. Regla general de protección

La protección de datos no deberá entenderse como una prohibición general de utilizar información sobre personas dentro del trabajo periodístico de Fundamedios. Su finalidad es asegurar que, cuando esa información sea tratada mediante inteligencia artificial, exista una decisión consciente respecto de **qué se utiliza, para qué, mediante qué herramienta, bajo qué protección y durante cuánto tiempo**.

El criterio aplicable durante todo el ciclo será:

utilizar la menor cantidad de datos personales necesaria para cumplir legítimamente la finalidad; protegerlos durante su tratamiento; limitar su acceso; verificar las inferencias producidas sobre las personas; y no conservarlos o reutilizarlos mediante IA más allá de lo razonablemente justificado.

3. Confidencialidad, fuentes periodísticas e información reservada

La protección de la información confidencial constituye una obligación independiente de la protección de datos personales. Una información puede requerir confidencialidad por su origen, por las condiciones bajo las cuales fue obtenida, por su relación con una investigación en curso, por una obligación legal o contractual, o porque su divulgación pueda generar riesgos para una persona o para Fundamedios.

La utilización de inteligencia artificial introduce riesgos particulares en este ámbito. Al incorporar información a una plataforma externa, cargar un documento, solicitar el análisis de una entrevista o incluir determinados antecedentes dentro de un prompt, el usuario puede estar comunicando información a un sistema operado por terceros, generando registros adicionales o permitiendo que determinados datos permanezcan almacenados fuera del repositorio en el que originalmente se encontraban.

Por esta razón, la confidencialidad deberá evaluarse **antes de proporcionar la información a la herramienta**, y no únicamente al momento de decidir qué se publica.

En el caso de las fuentes periodísticas, esta protección será especialmente rigurosa. La inteligencia artificial no deberá convertirse en un mecanismo mediante el cual la identidad de una fuente, o información que permita razonablemente identificarla, salga del círculo de confidencialidad dentro del cual fue obtenida.

3.1. Identificación de información confidencial

Se considerará confidencial, para efectos de este Manual, toda información respecto de la cual exista una expectativa razonable de acceso limitado, independientemente de que se encuentre o no formalmente marcada como “confidencial”.

Esta categoría podrá comprender documentos internos no publicados; investigaciones en curso; información entregada bajo reserva; comunicaciones internas; estrategias jurídicas; borradores sensibles; bases de datos restringidas; información sobre seguridad institucional; antecedentes de fuentes periodísticas; materiales obtenidos bajo compromisos de confidencialidad; y cualquier información cuya divulgación pueda perjudicar una investigación, una persona o los intereses legítimos de Fundamedios.

La ausencia de una etiqueta formal no elimina la obligación de proteger la información cuando, por su contenido o contexto, resulte evidente que no se encuentra destinada a circulación general.

Cuando exista duda razonable sobre su clasificación, deberá tratarse provisionalmente conforme al nivel superior de protección hasta determinar las condiciones adecuadas para su tratamiento.

3.2. Protección reforzada de las fuentes periodísticas

La identidad de una fuente periodística confidencial constituye información de protección reforzada.

Como regla general, **no deberá introducirse la identidad de una fuente confidencial en una herramienta de inteligencia artificial**, salvo que se utilice un entorno específicamente autorizado para ese nivel de información y exista una necesidad justificada para hacerlo.

Esta protección no se limita al nombre.

Deberán considerarse también aquellos elementos que, individualmente o combinados, puedan permitir identificar a la fuente: cargo, institución, área de trabajo, ubicación, participación en determinados acontecimientos, fechas de reuniones, datos de contacto, características personales, fragmentos de comunicaciones, metadatos u otros antecedentes.

Por ejemplo, sustituir el nombre de una fuente por “un funcionario de la institución X” puede resultar insuficiente si el resto del prompt señala su cargo exacto, la reunión en la que participó y la fecha en que obtuvo determinado documento.

La pregunta que deberá realizarse antes de utilizar IA será, por tanto, no solamente **“¿eliminé el nombre?”**, sino **“¿la herramienta necesita realmente conocer elementos que permitan identificar a esta persona?”**

3.3. Principio de separación de identidad y contenido

Cuando resulte necesario utilizar inteligencia artificial para trabajar con información proporcionada por una fuente protegida, deberá procurarse separar la **identidad de la fuente** del **contenido que necesita ser procesado**.

Así, por ejemplo, una entrevista podrá ser transcrita o analizada después de retirar, cuando sea posible y necesario, nombres, referencias personales, información contextual identificadora y otros elementos que no sean indispensables para la tarea.

Podrán emplearse identificadores internos tales como:

Fuente A — Fuente B — Testigo 01 — Documento R-03

o cualquier otro sistema que permita al equipo mantener la relación necesaria dentro del expediente protegido sin comunicarla innecesariamente a la herramienta.

La correspondencia entre el identificador y la identidad real deberá mantenerse fuera del material proporcionado a la IA cuando la herramienta no se encuentre autorizada para acceder a esa información.

3.4. No reconstrucción de identidades

Cuando se utilicen identificadores, anonimización o seudonimización para proteger a una fuente, **no deberá solicitarse posteriormente a la IA que intente determinar quién es esa persona** a partir del contexto disponible.

Tampoco deberán utilizarse otras herramientas, búsquedas automatizadas o cruces de bases de datos para revertir deliberadamente una medida de protección adoptada por Fundamedios, salvo que exista una finalidad periodística legítima distinta, debidamente evaluada y que no vulnere compromisos de confidencialidad asumidos por la organización.

Una medida de anonimización no deberá ser neutralizada por las propias capacidades analíticas de las herramientas utilizadas.

3.5. Información recibida bajo confidencialidad

Cuando Fundamedios reciba información bajo un compromiso expreso o razonablemente implícito de confidencialidad, las condiciones de ese compromiso deberán extenderse a su tratamiento mediante inteligencia artificial.

Antes de cargar el material deberá determinarse si la utilización de una herramienta externa resulta compatible con las condiciones bajo las cuales la información fue recibida.

La autorización para que Fundamedios conozca un documento **no implica necesariamente autorización para comunicarlo a un proveedor tecnológico externo**.

Cuando existan restricciones de este tipo, podrán utilizarse alternativas como trabajar únicamente con fragmentos no confidenciales, anonimizar previamente el material, realizar el procesamiento dentro de un entorno institucional específicamente habilitado o prescindir de IA para esa parte de la investigación.

Si ninguna alternativa permite preservar adecuadamente la confidencialidad, el documento deberá permanecer fuera del sistema de inteligencia artificial.

3.6. Investigaciones no publicadas

Las investigaciones en curso deberán considerarse información interna o confidencial según su naturaleza y nivel de riesgo.

Un expediente investigativo puede revelar no solamente información sobre terceros, sino también **qué está investigando Fundamedios, qué hipótesis examina, qué fuentes posee, qué documentos ha obtenido y qué líneas de investigación todavía no han sido publicadas**.

La exposición de esta información puede comprometer el trabajo periodístico incluso cuando ninguno de los documentos individuales sea secreto.

Por ello, los espacios de IA utilizados para investigaciones no publicadas deberán mantenerse dentro del entorno institucional y con acceso limitado al equipo que razonablemente necesite participar.

No deberán trasladarse conversaciones o expedientes a cuentas personales para continuar el trabajo fuera del entorno autorizado.

3.7. Borradores y productos previos a publicación

Los borradores generados mediante inteligencia artificial deberán conservar el nivel de protección correspondiente a la información que contienen.

Un texto todavía no publicado puede contener acusaciones preliminares, nombres que posteriormente serán eliminados, datos todavía sujetos a contraste, conclusiones provisionales o información que finalmente no será difundida.

Por ello, no deberá asumirse que un borrador puede circular libremente simplemente porque eventualmente se pretende publicar una versión del contenido.

La autorización para publicar la versión final no convierte retroactivamente en pública toda la información contenida en sus versiones anteriores.

3.8. Documentación jurídica y comunicaciones protegidas

Los expedientes jurídicos, estrategias procesales, consultas legales, informes de abogados y comunicaciones sometidas a deberes de confidencialidad deberán recibir un tratamiento compatible con la protección que les corresponda.

Cuando se utilice IA para apoyar análisis jurídicos deberá distinguirse entre documentos públicos —por ejemplo, determinadas sentencias o normas— y material interno cuya exposición pueda afectar una estrategia, una investigación o una relación protegida.

Antes de proporcionar este último tipo de información deberá verificarse que el entorno utilizado haya sido autorizado para su tratamiento.

Cuando el análisis pueda realizarse sustituyendo nombres, retirando antecedentes innecesarios o proporcionando únicamente la cuestión jurídica relevante, deberá preferirse esa alternativa.

3.9. Prompts como información institucional

Los prompts no deberán considerarse necesariamente instrucciones inocuas.

Una instrucción puede revelar una investigación confidencial aunque no incluya ningún documento adjunto. Preguntas como quién está siendo investigado, qué irregularidad se sospecha, qué información se intenta encontrar o qué estrategia se pretende seguir pueden revelar aspectos sustanciales del trabajo de Fundamedios.

Por ello, **la protección aplicable al contenido deberá extenderse al prompt que lo describe**.

Esto también significa que determinados prompts desarrollados por Fundamedios podrán constituir metodología o conocimiento institucional que deba conservarse dentro del entorno de la organización y no trasladarse sistemáticamente a cuentas personales.

3.10. Transcripciones y grabaciones

Las herramientas de inteligencia artificial utilizadas para transcribir reuniones, entrevistas, llamadas, audios o videos deberán ser evaluadas según la naturaleza de la conversación.

No deberá utilizarse automáticamente un servicio de transcripción externo para cualquier entrevista.

Antes de hacerlo deberá considerarse:

- si las personas conocen o deben conocer la existencia de la grabación conforme a las reglas aplicables;
- si existen fuentes protegidas;
- si la conversación contiene información confidencial;
- dónde será procesado el audio;
- cuánto tiempo conservará el proveedor la grabación;
- si el contenido podrá ser reutilizado para otros fines; y
- si existe una alternativa institucional más segura.

Cuando una grabación contenga información de riesgo alto, deberá utilizarse únicamente una herramienta autorizada para ese nivel de tratamiento o realizarse el procesamiento mediante un mecanismo alternativo.

3.11. Confidencialidad en espacios compartidos de IA

Los entornos institucionales podrán permitir proyectos compartidos, asistentes especializados, bibliotecas documentales o bases de conocimiento comunes. Su carácter institucional no significa que toda la información cargada en ellos pueda ser conocida por todas las personas de la organización.

Antes de incorporar material confidencial a un espacio compartido deberá comprobarse quién tiene acceso efectivo a ese espacio y qué capacidades posee.

Los permisos deberán establecerse según necesidad.

Cuando un proyecto contenga fuentes protegidas o investigaciones sensibles, deberá preferirse un espacio separado antes que confiar únicamente en que los usuarios no consultarán información que no les corresponde.

La arquitectura tecnológica deberá acompañar la regla organizativa de confidencialidad.

3.12. Divulgación accidental a través de resultados generados

La protección de información confidencial no termina después de cargarla en una herramienta.

Cuando una IA tenga acceso a repositorios, memorias o conversaciones previas, existe la posibilidad de que incorpore información proveniente de ellos en una respuesta posterior.

Por esta razón, los sistemas institucionales deberán configurarse de manera que reduzcan el riesgo de **contaminación entre proyectos o contextos**.

Una persona que tenga autorización para utilizar un asistente general no debería recibir, por esa sola circunstancia, información proveniente de una investigación restringida.

Cuando la plataforma no permita establecer separaciones adecuadas, la información de mayor sensibilidad deberá mantenerse fuera de ese entorno.

3.13. Prohibiciones específicas

Con independencia de las demás reglas establecidas en este Manual, quedará prohibido:

I. introducir la identidad de una fuente confidencial en una herramienta no autorizada para ese nivel de información;

II. cargar documentos reservados en cuentas personales de inteligencia artificial;

III. proporcionar a una herramienta credenciales, contraseñas, claves privadas, tokens o mecanismos de autenticación;

IV. utilizar IA para intentar descubrir la identidad de una fuente cuya confidencialidad Fundamedios se haya comprometido a proteger;

V. incorporar información confidencial a una memoria, asistente o base de conocimiento de acceso general;

VI. compartir externamente conversaciones de IA que contengan información protegida sin realizar previamente la depuración correspondiente;

VII. utilizar herramientas gratuitas o no evaluadas para procesar información restringida por el solo hecho de que permitan realizar técnicamente la tarea; y

VIII. copiar información institucional protegida a cuentas personales para evitar limitaciones, permisos o controles del entorno institucional.

3.14. Divulgación interna mínima

La confidencialidad también opera dentro de la organización.

La información protegida deberá ser conocida únicamente por las personas cuya intervención sea razonablemente necesaria para desarrollar, verificar, revisar, proteger o aprobar el trabajo correspondiente.

La incorporación de IA no deberá ampliar innecesariamente ese círculo.

Cuando una persona necesite asistencia técnica para resolver un problema relacionado con una herramienta, deberá procurar describir el problema sin revelar información confidencial que no sea necesaria para recibir soporte.

Del mismo modo, los registros técnicos y reportes de incidentes deberán contener suficiente información para investigar lo ocurrido, pero no reproducir innecesariamente datos protegidos.

3.15. Publicación y levantamiento de la confidencialidad

La publicación de determinada información puede modificar su nivel de confidencialidad, pero deberá analizarse qué elementos fueron efectivamente hechos públicos.

Si Fundamedios publica una investigación basada en documentos confidenciales, ello no significa necesariamente que los documentos completos, la identidad de las fuentes o todos los antecedentes recopilados pasen a ser públicos.

La versión publicada y el expediente interno deberán continuar tratándose como elementos diferentes.

Cuando una fuente autorice posteriormente revelar su identidad o determinado documento deje de estar sujeto a reserva, deberá documentarse suficientemente el cambio antes de modificar los controles aplicables.

3.16. Incidentes de confidencialidad

Cuando exista la posibilidad de que información confidencial haya sido introducida indebidamente en una herramienta, compartida con usuarios no autorizados o expuesta mediante un resultado generado, deberá actuarse de inmediato.

Las primeras medidas deberán orientarse a **contener la exposición**, no a determinar culpabilidades.

Según el caso, podrá ser necesario restringir accesos, eliminar archivos o conversaciones cuando técnicamente sea posible, suspender temporalmente una integración, cambiar credenciales, preservar registros del incidente y determinar qué información pudo haber quedado comprometida.

Si el incidente involucra una fuente periodística protegida o puede generar un riesgo para la seguridad de una persona, deberá escalarse con carácter prioritario.

La determinación de responsabilidades se realizará posteriormente conforme al procedimiento establecido en este Manual.

3.17. Regla de máxima precaución para fuentes protegidas

Cuando exista una duda razonable sobre si determinada utilización de inteligencia artificial puede comprometer directa o indirectamente la identidad de una fuente confidencial, deberá preferirse **no proporcionar la información al sistema** hasta contar con un mecanismo que permita procesarla de manera segura.

En esta materia, la utilidad o conveniencia de la herramienta no prevalecerá sobre la protección de la fuente.

La regla operativa será:

La IA podrá ayudar a trabajar con la información de una fuente; nunca deberá convertirse en la razón por la que esa fuente deje de estar protegida.

4. Almacenamiento, conservación y eliminación de la información

La utilización de inteligencia artificial puede multiplicar rápidamente la cantidad de información que genera una investigación. A los documentos originales pueden añadirse transcripciones, extracciones, tablas, resúmenes, bases estructuradas, versiones de trabajo, prompts, resultados de búsqueda, borradores y materiales de verificación. Si estos elementos no se administran adecuadamente, una misma investigación puede terminar distribuida entre conversaciones de IA, computadoras personales, servicios de almacenamiento, correos electrónicos y diferentes versiones de un mismo archivo.

Fundamedios deberá procurar que la información que tenga valor para el trabajo institucional permanezca en **repositorios identificables, organizados y administrados por la organización**. El objetivo no será conservar todo lo que produzca una herramienta, sino preservar aquello que permita desarrollar, comprobar, continuar o reconstruir razonablemente el trabajo realizado.

El almacenamiento deberá responder simultáneamente a cuatro criterios: **utilidad, trazabilidad, seguridad y necesidad de conservación**.

4.1. Repositorio institucional como lugar principal de conservación

Los documentos y materiales que deban conservarse como parte de una investigación o producto deberán almacenarse en los sistemas institucionales definidos por Fundamedios.

Las conversaciones mantenidas con una herramienta de inteligencia artificial no deberán utilizarse como único repositorio de una investigación.

Una conversación puede desaparecer, quedar vinculada a un usuario determinado, cambiar por modificaciones de la plataforma o resultar difícil de consultar posteriormente. Además, puede contener simultáneamente información válida, hipótesis, errores, instrucciones y resultados descartados.

Por ello, aquello que resulte relevante deberá trasladarse o vincularse al expediente, proyecto o repositorio institucional correspondiente.

Esto podrá comprender, según el caso, fuentes principales, documentos obtenidos, bases de datos utilizadas, registros de verificación, materiales audiovisuales relevantes, productos intermedios que sea necesario preservar y la versión finalmente aprobada.

4.2. Estructura y organización

No será necesario imponer una estructura documental idéntica para todos los proyectos. La organización de una alerta diaria puede ser mucho más sencilla que la de una investigación desarrollada durante varios meses.

Sin embargo, los repositorios deberán permitir encontrar razonablemente la información y reconocer qué función cumple cada elemento.

En investigaciones de cierta complejidad podrá resultar útil separar, por ejemplo:

01 — Fuentes y documentos originales 02 — Información recopilada y bases de trabajo 03 — Verificación y contraste 04 — Material restringido 05 — Análisis y documentos de trabajo 06 — Producción y versiones editoriales 07 — Producto final

Esta estructura es referencial. Los equipos podrán utilizar otros sistemas siempre que mantengan un orden suficiente y respeten las reglas de acceso y protección establecidas en este Manual.

Lo importante es evitar repositorios en los cuales resulte imposible distinguir una fuente original de un resumen producido por IA, un documento vigente de una versión anterior o un hallazgo validado de información posteriormente descartada.

4.3. Conservación de originales

Cuando una investigación dependa de un documento, archivo, fotografía, audio, video, base de datos u otra fuente relevante, deberá procurarse conservar el original obtenido o una copia suficientemente fiel cuando resulte necesario y jurídicamente procedente.

Los resultados transformados mediante IA no deberán sustituir al material original.

Una transcripción no reemplaza necesariamente al audio; un resumen no reemplaza al documento; una tabla de datos extraídos no reemplaza a la base de la cual fueron obtenidos; y una descripción generada de una imagen no reemplaza a la imagen.

La conservación del original permite posteriormente comprobar errores de extracción, interpretación o generación.

4.4. Nomenclatura y versiones

En proyectos que produzcan múltiples documentos deberá utilizarse una nomenclatura que permita reconocer razonablemente su contenido y estado.

Deberán evitarse, en la medida de lo posible, cadenas de archivos denominados simplemente “final”, “final 2”, “final nuevo” o equivalentes que impidan determinar cuál fue efectivamente aprobado.

Cuando existan modificaciones sustanciales, podrá incorporarse fecha, número de versión, estado u otra identificación útil.

No será necesario registrar cada modificación menor. El control de versiones deberá concentrarse en aquellos cambios que tengan importancia para reconstruir el proceso o evitar que una versión superada vuelva a utilizarse.

4.5. Qué debe conservarse de las interacciones con IA

No toda conversación, prompt o respuesta producida mediante inteligencia artificial deberá archivarse.

La conservación será necesaria cuando la interacción tenga una función relevante para comprender o reconstruir el trabajo. Esto podrá ocurrir, entre otros casos, cuando:

- el prompt constituya una metodología institucional reutilizable;
- la IA haya realizado una extracción o análisis que sea sustancial para la investigación;
- resulte necesario conservar evidencia de cómo se obtuvo determinado resultado;
- la conversación documente una decisión o proceso relevante;
- el resultado haya sido utilizado como insumo importante para un producto; o
- su conservación sea necesaria para revisar posteriormente un incidente, error o cuestionamiento.

Las consultas rutinarias, pruebas, reformulaciones sin valor posterior y demás interacciones que no tengan relevancia institucional no necesitarán incorporarse al expediente.

La trazabilidad no exige almacenar cada palabra intercambiada con una IA.

4.6. Información descartada

La información que haya sido descartada por falsa, incorrecta o insuficientemente sustentada deberá manejarse de manera que no pueda confundirse posteriormente con información validada.

Dependiendo de su relevancia, podrá eliminarse o conservarse separadamente cuando resulte necesario documentar por qué se descartó una línea de investigación.

Si se conserva, deberá quedar claramente identificada como **DESCARTADA**, **NO VERIFICADA**, **SUSTITUIDA** o mediante otra denominación inequívoca.

Esta regla adquiere especial importancia en repositorios conectados a sistemas de inteligencia artificial. La información descartada no deberá permanecer dentro de una base de conocimiento activa de forma que pueda ser recuperada posteriormente como si se tratara de conocimiento institucional válido.

4.7. Plazos de conservación

Fundamedios deberá establecer períodos de conservación razonables de acuerdo con la naturaleza de la información.

No todos los materiales deberán conservarse durante el mismo tiempo.

Para determinar el período aplicable podrán considerarse:

a. la finalidad para la cual se obtuvo la información;

b. su valor periodístico, documental, histórico o institucional;

c. la posibilidad de que sea necesaria para respaldar una publicación;

d. obligaciones jurídicas o contractuales;

e. la existencia de procedimientos administrativos, judiciales o controversias relacionadas;

f. la presencia de datos personales o información confidencial;

g. los riesgos derivados de continuar conservándola; y

h. la posibilidad de mantener su utilidad mediante anonimización.

Cuando exista una controversia, reclamación, procedimiento judicial o una expectativa razonable de que determinado material pueda resultar necesario para la defensa de Fundamedios, no deberá eliminarse información relevante sin considerar previamente esa circunstancia.

4.8. Revisión periódica de la información almacenada

Los repositorios de IA y las bases de conocimiento institucional deberán ser objeto de revisiones periódicas.

El propósito será detectar información que:

- ya no resulte necesaria;
- haya quedado desactualizada;
- haya sido sustituida;
- contenga errores conocidos;
- conserve permisos de acceso innecesariamente amplios;
- incluya datos personales que ya no resulte justificado mantener; o
- haya perdido la finalidad que justificaba su incorporación a una memoria activa de IA.

Esta revisión deberá concentrarse especialmente en los repositorios que la inteligencia artificial pueda consultar automáticamente, porque el riesgo no consiste únicamente en conservar información innecesaria, sino en que **vuelva a ser utilizada para generar nuevos contenidos**.

4.9. Archivo y memoria activa no son lo mismo

Cuando determinada información deba conservarse por razones documentales, jurídicas, históricas o periodísticas, pero ya no resulte necesario que sea utilizada regularmente por una herramienta de IA, deberá evaluarse su traslado a un archivo institucional separado de la memoria activa.

De esta forma, Fundamedios podrá conservar el expediente sin mantenerlo permanentemente disponible para nuevos procesos automatizados.

Esta separación será particularmente útil para investigaciones cerradas, expedientes antiguos, información sensible y grandes conjuntos de datos personales.

La memoria activa deberá contener aquello que tenga utilidad razonable para el trabajo actual o recurrente. El archivo podrá responder a necesidades de conservación más prolongadas.

4.10. Eliminación segura

Cuando corresponda eliminar información, la eliminación deberá comprender, en la medida en que resulte técnicamente posible y jurídicamente exigible, los lugares institucionales en los cuales se encuentre almacenada.

No deberá asumirse que borrar el archivo original elimina necesariamente las copias, versiones derivadas o registros existentes en otros sistemas.

Según el caso, deberá considerarse su presencia en:

repositorios institucionales — herramientas de IA — proyectos compartidos — bases de conocimiento — dispositivos — sistemas de respaldo — servicios externos autorizados.

Las posibilidades técnicas de eliminación pueden variar entre plataformas. Por esta razón, las capacidades de conservación y supresión de información deberán formar parte de la evaluación previa de las herramientas autorizadas.

Cuando una plataforma no permita eliminar razonablemente información que por su naturaleza pueda requerir supresión posterior, esta limitación deberá ser considerada antes de autorizar su utilización para ese tipo de datos.

4.11. Salida o cambio de funciones de una persona

Cuando una persona deje Fundamedios o cambie de funciones, deberá garantizarse que la información institucional necesaria para continuar su trabajo permanezca bajo control de la organización.

Antes de retirar sus accesos deberá procurarse, cuando corresponda:

1. identificar los proyectos institucionales bajo su responsabilidad;
2. transferir los documentos y materiales que deban conservarse;
3. comprobar que la información relevante no permanezca únicamente en espacios personales;
4. reasignar la administración de proyectos, asistentes o repositorios de IA;
5. retirar o modificar los permisos que ya no sean necesarios; y
6. revocar las credenciales institucionales correspondientes.

La terminación de la relación con Fundamedios no autorizará a conservar copias personales de información confidencial, restringida o de otra información institucional cuya retención no corresponda.

4.12. Respaldo y recuperación

La información cuya pérdida pueda afectar significativamente el trabajo de Fundamedios deberá contar con mecanismos razonables de respaldo y recuperación.

La dependencia de una plataforma de inteligencia artificial no deberá convertirse en un punto único de pérdida de información.

Los documentos centrales, fuentes y productos institucionales deberán permanecer disponibles mediante los sistemas de conservación definidos por Fundamedios aun cuando una herramienta de IA deje de funcionar, cambie sus condiciones, suspenda una cuenta o elimine una determinada funcionalidad.

La continuidad del conocimiento institucional deberá depender de **Fundamedios y sus sistemas de información**, no de la permanencia indefinida de un proveedor específico de inteligencia artificial.

4.13. Eliminación de información de cuentas personales

Si se detecta que información institucional fue procesada o almacenada indebidamente en una cuenta personal, deberá evaluarse su traslado al entorno institucional cuando corresponda conservarla y su posterior eliminación de la cuenta personal.

Cuando la información sea confidencial, restringida o contenga datos personales relevantes, deberá además determinarse si el hecho constituye un incidente que requiera registro, evaluación o medidas adicionales.

La simple transferencia de una copia al entorno institucional no subsana por sí misma la permanencia de información protegida en un espacio no autorizado.

4.14. Regla de conservación mínima necesaria

El almacenamiento institucional no deberá convertirse en acumulación indefinida.

La capacidad de conservar grandes volúmenes de información a bajo costo no elimina los riesgos derivados de mantenerlos.

Cada dato conservado puede generar una utilidad futura, pero también representa una superficie adicional de exposición, acceso indebido, utilización fuera de contexto o generación de resultados incorrectos.

Por ello, Fundamedios aplicará como criterio general:

conservar lo que sea necesario para proteger, comprobar, continuar o documentar el trabajo institucional; archivar aquello que deba preservarse pero no necesite permanecer activo; y eliminar aquello cuya conservación haya dejado de estar razonablemente justificada.

4.15. Trazabilidad después de la publicación

La publicación de un producto no deberá provocar la eliminación automática de los elementos necesarios para demostrar posteriormente cómo fueron comprobadas sus afirmaciones centrales.

En función de la naturaleza y riesgo del contenido, Fundamedios deberá conservar durante un período razonable las fuentes y registros suficientes para responder preguntas, realizar correcciones, atender solicitudes de aclaración, defender la integridad de la publicación o reconstruir el proceso seguido.

Esto adquiere especial importancia cuando haya intervenido inteligencia artificial.

Si posteriormente se cuestiona una publicación, la organización deberá poder distinguir **qué información provino de una fuente, qué procesamiento realizó la IA, qué verificó una persona y qué versión fue finalmente autorizada para publicación**.

La trazabilidad constituye, por tanto, no solo una herramienta de organización, sino también un mecanismo de protección institucional.

5. Ciberseguridad y seguridad en el uso de herramientas de inteligencia artificial

La seguridad de la información utilizada mediante inteligencia artificial no depende únicamente de la confidencialidad de los documentos o de la protección de los datos personales. Depende también de la seguridad de las cuentas, dispositivos, integraciones, repositorios y servicios tecnológicos que permiten acceder a esa información.

Una herramienta de IA puede tener acceso no solamente a aquello que una persona escribe directamente en un prompt. Dependiendo de su configuración, puede acceder a archivos cargados, memorias, espacios compartidos, repositorios documentales, servicios en la nube, correos electrónicos, calendarios, navegadores, bases de datos u otros sistemas conectados. Mientras mayor sea esta capacidad de acceso, mayor será también el impacto potencial de una cuenta comprometida, una configuración incorrecta o una instrucción maliciosa.

Por esta razón, Fundamedios aplicará un criterio de **seguridad proporcional al acceso y al riesgo**. Una herramienta destinada únicamente a trabajar con información pública no requerirá necesariamente los mismos controles que un sistema conectado a investigaciones no publicadas, documentos internos o bases institucionales.

La seguridad deberá diseñarse procurando que un incidente aislado no permita acceder automáticamente a toda la información de la organización.

5.1. Seguridad de las cuentas institucionales

Las cuentas institucionales utilizadas para acceder a sistemas de inteligencia artificial deberán estar protegidas mediante mecanismos adecuados de autenticación.

Siempre que la plataforma lo permita, deberá utilizarse **autenticación multifactor** para cuentas institucionales y, de manera obligatoria, para aquellas con privilegios administrativos o acceso a información confidencial o restringida.

Las credenciales serán individuales y no deberán compartirse entre usuarios. Cuando varias personas necesiten trabajar sobre una misma investigación, deberá utilizarse la funcionalidad de equipos, proyectos, espacios compartidos o permisos proporcionada por la plataforma, en lugar de compartir una contraseña común.

Las contraseñas deberán cumplir las políticas de seguridad institucionales y no reutilizarse entre servicios cuando ello incremente el riesgo de compromiso.

Deberá prestarse particular atención a las cuentas con capacidad para administrar usuarios, modificar configuraciones, acceder a facturación, incorporar integraciones, consultar registros o modificar bases de conocimiento. Estas cuentas poseen un nivel de riesgo superior y sus privilegios deberán limitarse a las personas que realmente los necesiten.

5.2. Principio de mínimo privilegio

Toda herramienta, usuario, integración o servicio conectado deberá disponer únicamente de los permisos necesarios para cumplir su función.

Cuando una aplicación solicite acceso a un repositorio completo, correo electrónico, almacenamiento institucional u otro sistema, no deberá concederse automáticamente ese acceso por comodidad.

Antes de autorizar una integración deberá preguntarse:

¿Qué necesita consultar? ¿Qué necesita modificar? ¿Durante cuánto tiempo necesita acceso? ¿Puede limitarse a una carpeta, proyecto o conjunto de información específico?

Cuando técnicamente sea posible, deberán preferirse permisos de lectura frente a permisos de modificación si la tarea no requiere alterar información.

Un asistente destinado a resumir documentos, por ejemplo, no debería recibir capacidad para eliminar archivos si esa función no resulta necesaria.

5.3. Integraciones, conectores y herramientas externas

Las capacidades de una plataforma de inteligencia artificial pueden ampliarse mediante extensiones, complementos, conectores, APIs, aplicaciones externas y otros servicios.

Cada integración amplía también el perímetro tecnológico dentro del cual puede circular la información institucional.

Antes de habilitar una integración que pueda acceder a información de Fundamedios deberá evaluarse, de acuerdo con su riesgo:

- qué información podrá consultar;
- si podrá modificar o eliminar contenido;
- quién opera el servicio;
- qué información recibirá el tercero;
- qué permisos solicita;
- cómo se autentica;
- durante cuánto tiempo conserva los datos;
- si existen subproveedores;
- cómo puede revocarse el acceso; y
- qué ocurrirá con la información cuando la integración deje de utilizarse.

La autorización de una plataforma de IA **no implica automáticamente la autorización de todas las aplicaciones que puedan conectarse a ella**.

5.4. Acceso a repositorios y bases de conocimiento

Cuando una herramienta de inteligencia artificial tenga acceso directo a repositorios institucionales, deberá evitarse conceder acceso indiscriminado a toda la documentación de Fundamedios.

Siempre que la tecnología lo permita, los repositorios deberán segmentarse por proyecto, finalidad, nivel de confidencialidad o grupo autorizado.

Una herramienta destinada a apoyar la producción de notas públicas puede requerir acceso a investigaciones publicadas, manuales editoriales y fuentes institucionales, pero no necesariamente a información sobre fuentes confidenciales, expedientes jurídicos o investigaciones restringidas.

El principio aplicable será:

el hecho de que una IA pueda técnicamente consultar toda la información institucional no significa que deba permitírsele hacerlo.

La arquitectura de acceso deberá procurar que un error en un proyecto no exponga información perteneciente a otro.

5.5. Dispositivos utilizados para acceder a IA institucional

El acceso a herramientas institucionales deberá realizarse desde dispositivos que cuenten con medidas razonables de seguridad.

Los equipos deberán mantenerse actualizados y protegidos mediante los mecanismos establecidos por Fundamedios. Cuando se utilicen dispositivos personales para actividades institucionales, deberán respetarse las condiciones de seguridad que determine la organización.

No deberá mantenerse información institucional sensible descargada localmente durante más tiempo del necesario.

Especial cuidado deberá tenerse con computadoras compartidas, dispositivos públicos o equipos sobre los cuales Fundamedios no pueda establecer condiciones mínimas de seguridad.

Las sesiones institucionales no deberán permanecer abiertas en dispositivos a los que puedan acceder personas no autorizadas.

5.6. Archivos recibidos de terceros

Las herramientas de inteligencia artificial permiten procesar rápidamente documentos recibidos por correo electrónico, mensajería, redes sociales, fuentes o terceros. Esta facilidad no elimina los riesgos tecnológicos asociados con archivos de origen desconocido.

Un archivo deberá tratarse de acuerdo con su procedencia y nivel de riesgo antes de incorporarlo a sistemas institucionales.

Cuando existan razones para sospechar que un archivo puede ser malicioso, deberá evitarse su apertura o procesamiento ordinario hasta realizar los controles correspondientes.

Esta regla adquiere especial importancia en investigaciones periodísticas, donde una persona puede recibir deliberadamente archivos diseñados para comprometer un dispositivo, obtener credenciales o acceder a información sobre una investigación.

5.7. Enlaces y contenidos externos

La IA puede localizar y recomendar enlaces durante una investigación. El hecho de que un enlace haya sido proporcionado por una herramienta institucional no garantiza que el sitio sea seguro.

Antes de proporcionar credenciales, descargar archivos o conceder permisos desde un enlace encontrado mediante IA deberá verificarse razonablemente el destino.

Deberá prestarse especial atención a sitios que imiten plataformas conocidas, páginas que soliciten volver a iniciar sesión, dominios similares a los legítimos, descargas inesperadas o solicitudes de autorización para acceder a cuentas institucionales.

La inteligencia artificial puede ayudar a encontrar información en Internet, pero no deberá convertirse en una vía automática para confiar en cualquier recurso que aparezca dentro de sus resultados.

5.8. Instrucciones maliciosas y

prompt injection

Los sistemas de inteligencia artificial que consultan páginas web, documentos, correos electrónicos, archivos u otras fuentes pueden encontrar dentro de esos contenidos instrucciones dirigidas al propio sistema.

Estas instrucciones pueden intentar modificar su comportamiento, hacer que ignore las indicaciones proporcionadas por Fundamedios, revelar información a la que tenga acceso, ejecutar acciones no solicitadas o

utilizar herramientas conectadas de una manera diferente a la prevista. Este tipo de riesgo puede presentarse mediante técnicas conocidas como *prompt injection* o inyección de instrucciones.

El contenido de una fuente externa deberá tratarse, por tanto, **como información que debe analizarse y no como instrucciones que la IA deba obedecer**.

Cuando un asistente tenga acceso simultáneamente a Internet y a información institucional protegida, deberá aplicarse una precaución reforzada.

Los sistemas no deberán estar configurados de forma que una instrucción encontrada dentro de una página web o documento pueda provocar automáticamente la divulgación, modificación o transferencia de información institucional.

5.9. Acciones ejecutadas por agentes de inteligencia artificial

Cuando Fundamedios utilice sistemas capaces no solamente de generar texto, sino también de ejecutar acciones —por ejemplo, enviar comunicaciones, modificar documentos, publicar contenido, consultar bases, crear registros o interactuar con otros sistemas— deberán establecerse controles adicionales.

La autorización para **proponer una acción** no deberá confundirse con la autorización para **ejecutarla**.

Las acciones que puedan producir efectos externos relevantes deberán requerir revisión o confirmación humana antes de su ejecución, particularmente cuando impliquen:

a) publicar contenido; b) enviar comunicaciones externas; c) eliminar o modificar información institucional; d) compartir archivos; e) conceder accesos;

f) realizar cambios en bases de datos; g) **comunicarse en nombre de Fundamedios;** o **h) ejecutar cualquier acción difícil de revertir.**

Las funciones autónomas deberán habilitarse únicamente cuando exista una finalidad clara y el riesgo de la acción pueda ser razonablemente controlado.

5.10. Prevención de extracción de información

Una herramienta conectada a información institucional puede ser objeto de intentos deliberados de obtener contenido al cual el solicitante no debería tener acceso.

Por ejemplo, una persona podría formular instrucciones diseñadas para conseguir que el sistema revele documentos internos, reproduzca conversaciones anteriores, exponga información contenida en su memoria o describa instrucciones institucionales reservadas.

Los sistemas deberán configurarse, cuando sea técnicamente posible, para limitar este tipo de extracción.

Sin embargo, las restricciones del modelo no deberán constituir la única barrera de seguridad.

La protección principal deberá provenir de **los permisos de acceso y de la separación de información**. Si una persona nunca debería conocer determinado documento, la opción más segura será que el asistente utilizado por esa persona tampoco tenga acceso a él.

5.11. Descarga y exportación de conversaciones

Las plataformas de inteligencia artificial pueden permitir compartir conversaciones mediante enlaces, exportarlas, descargarlas o transferirlas a otros usuarios.

Estas funciones deberán utilizarse con el mismo cuidado que cualquier otro mecanismo de transferencia de información.

Antes de compartir una conversación deberá revisarse su contenido completo y no únicamente la última respuesta.

Una conversación puede contener en mensajes anteriores datos personales, documentos, hipótesis, nombres de fuentes, instrucciones internas o información descartada que no debería acompañar al contenido que se pretende compartir.

Cuando únicamente sea necesario comunicar un resultado, deberá preferirse trasladar el resultado depurado antes que compartir automáticamente todo el historial de conversación.

5.12. Configuraciones de privacidad y entrenamiento

Las herramientas institucionales deberán configurarse, cuando la plataforma ofrezca estas opciones, para reducir la utilización innecesaria de información de Fundamedios para finalidades diferentes de la prestación del servicio.

Deberán revisarse particularmente las configuraciones relacionadas con:

- utilización de conversaciones para entrenamiento o mejora de modelos;
- memoria entre conversaciones;
- historial;
- retención de datos;
- acceso de terceros;
- publicación o compartición de conversaciones;
- integraciones automáticas; y
- permisos de los administradores y usuarios.

Los cambios relevantes en estas configuraciones deberán ser administrados institucionalmente y no quedar exclusivamente a criterio de cada usuario cuando puedan afectar la seguridad de la organización.

5.13. Actualizaciones y cambios de funcionalidad

Las plataformas de inteligencia artificial cambian con frecuencia. Una actualización puede incorporar memoria, navegación web, capacidad de ejecutar código, acceso a aplicaciones, nuevos conectores o mecanismos diferentes de tratamiento de información.

Una funcionalidad nueva no deberá considerarse automáticamente autorizada porque aparezca dentro de una herramienta que Fundamedios ya utiliza.

Cuando el cambio amplíe sustancialmente aquello que el sistema puede consultar, conservar, transmitir o ejecutar, deberá evaluarse antes de incorporarlo a actividades institucionales sensibles.

Las herramientas autorizadas deberán revisarse periódicamente para comprobar que las condiciones bajo las cuales fueron inicialmente aprobadas continúan siendo válidas.

5.14. Registro y monitoreo de accesos

Cuando la plataforma permita consultar registros de actividad, Fundamedios podrá utilizarlos para administrar la seguridad de sus entornos institucionales.

Los registros podrán resultar útiles para conocer accesos, cambios de configuración, incorporación de usuarios, utilización de integraciones u otras actuaciones relevantes.

El monitoreo deberá realizarse de forma proporcional y con finalidades de seguridad, administración y cumplimiento.

En sistemas que contengan información de riesgo alto deberá procurarse conservar suficiente trazabilidad para determinar quién accedió o modificó información relevante cuando sea necesario investigar un incidente.

5.15. Detección de comportamientos anómalos

Cualquier persona que utilice una herramienta institucional deberá prestar atención a comportamientos que puedan indicar un problema de seguridad.

Entre otros:

- aparición inesperada de información perteneciente a otro proyecto;
- respuestas que contengan datos que el sistema aparentemente no debería conocer;
- cambios no autorizados en documentos o configuraciones;
- solicitudes inusuales de credenciales;
- accesos o notificaciones de inicio de sesión no reconocidos;
- nuevas integraciones que el usuario no haya habilitado;
- pérdida repentina de permisos;
- archivos que aparezcan modificados sin explicación; o
- respuestas del sistema que indiquen haber ejecutado acciones no solicitadas.

Cuando exista una duda razonable, deberá evitarse continuar proporcionando información sensible hasta determinar qué ocurrió.

5.16. Incidentes de seguridad relacionados con IA

Se considerará incidente de seguridad relacionado con inteligencia artificial cualquier evento que pueda comprometer la confidencialidad, integridad, disponibilidad o control de información institucional mediante una herramienta de IA o sistema relacionado.

Puede existir un incidente aunque todavía no se haya comprobado un daño.

Podrán constituir incidentes, entre otros, el acceso no autorizado a una cuenta; la carga de información restringida en una herramienta no autorizada; la exposición de una conversación; la pérdida de un dispositivo con una sesión institucional abierta; la utilización indebida de una integración; la revelación de información perteneciente a otro proyecto; o la ejecución de una acción automatizada no autorizada.

La persona que detecte el incidente deberá comunicarlo **tan pronto como sea posible** mediante el mecanismo institucional establecido.

5.17. Respuesta inicial ante un incidente

La respuesta deberá concentrarse inicialmente en contener el problema y preservar la información necesaria para comprenderlo.

Dependiendo de las circunstancias podrán adoptarse, entre otras, las siguientes medidas:

Contener. Suspender accesos, cerrar sesiones, revocar permisos, deshabilitar integraciones o impedir temporalmente el uso de la herramienta.

Proteger. Cambiar credenciales, activar o reforzar autenticación y restringir el acceso a repositorios relacionados.

Preservar evidencia. Conservar registros, capturas, fechas, mensajes y demás información necesaria para reconstruir lo ocurrido.

Determinar alcance. Identificar qué información pudo verse comprometida, durante cuánto tiempo y quién pudo tener acceso.

Corregir. Eliminar información cuando resulte posible, modificar configuraciones, corregir permisos o sustituir mecanismos vulnerables.

Evaluar obligaciones. Determinar si el incidente activa deberes jurídicos, contractuales, de protección de datos, comunicación a personas afectadas o cualquier otra actuación.

Aprender. Una vez contenido el incidente, identificar qué control falló y qué modificación puede impedir su repetición.

5.18. Protección de fuentes ante incidentes tecnológicos

Cuando un incidente pueda involucrar la identidad o información que permita identificar a una fuente confidencial, deberá recibir prioridad especial.

La evaluación deberá considerar no solamente si el nombre de la fuente fue expuesto, sino si la combinación de documentos, fechas, comunicaciones, metadatos u otros antecedentes puede permitir identificarla.

Cuando exista un riesgo razonable para la fuente, deberán adoptarse inmediatamente las medidas de protección correspondientes y escalarse el incidente a las personas responsables.

5.19. Seguridad frente a errores humanos

Una parte importante de los incidentes puede producirse sin que exista un ataque tecnológico.

Seleccionar el archivo equivocado, copiar información en una conversación incorrecta, compartir un enlace con permisos abiertos, confundir una cuenta personal con la institucional o conceder acceso excesivo a una integración pueden producir efectos similares a una vulneración externa.

Por ello, la seguridad del uso de IA deberá combinar controles técnicos con prácticas sencillas de trabajo.

Antes de cargar información sensible deberá comprobarse:

qué archivo se está proporcionando — a qué herramienta — dentro de qué cuenta — en qué proyecto — y quién puede acceder posteriormente.

Cuando la operación sea especialmente sensible, una comprobación adicional antes de ejecutarla puede evitar un incidente difícil de revertir.

5.20. Revisión periódica de accesos y permisos

Fundamedios deberá revisar periódicamente los accesos a sus principales herramientas de inteligencia artificial y repositorios asociados.

La revisión deberá permitir retirar cuentas de personas que ya no participan en la organización, eliminar permisos que dejaron de ser necesarios, identificar administradores innecesarios, revocar integraciones antiguas y comprobar que los espacios sensibles continúan restringidos.

También deberá realizarse una revisión cuando una persona cambie significativamente de funciones o deje de participar en un proyecto de acceso restringido.

Los permisos no deberán permanecer indefinidamente únicamente porque alguna vez fueron necesarios.

5.21. Continuidad y dependencia tecnológica

Fundamedios deberá evitar que una herramienta de inteligencia artificial se convierta en un punto único de dependencia para información crítica.

Los documentos esenciales, metodologías, fuentes y productos que deban conservarse deberán permanecer accesibles institucionalmente aun cuando un proveedor suspenda el servicio, modifique sus condiciones, elimine una funcionalidad o Fundamedios decida dejar de utilizarlo.

Cuando una herramienta cumpla una función especialmente importante deberá considerarse cómo recuperar la información necesaria y continuar el trabajo en caso de indisponibilidad.

La IA es una herramienta dentro de la infraestructura institucional; **la continuidad de Fundamedios no deberá depender de una plataforma determinada**.

5.22. Cultura de seguridad

La seguridad no deberá entenderse como responsabilidad exclusiva de quien administre los sistemas tecnológicos.

Cada persona que utilice inteligencia artificial dentro de Fundamedios participa en la protección de la información a la que tiene acceso.

Esto implica utilizar las herramientas autorizadas, proteger las credenciales, respetar los permisos, evitar transferencias innecesarias, reconocer situaciones de riesgo y comunicar oportunamente los incidentes.

Un error comunicado rápidamente puede ser contenido. Un error ocultado o ignorado puede convertirse en una exposición mucho mayor.

La comunicación de buena fe de un posible incidente deberá, por tanto, ser promovida como parte del sistema de prevención.

5.23. Regla general de seguridad

La incorporación de nuevas capacidades de inteligencia artificial deberá acompañarse de controles proporcionales a aquello que esas capacidades permiten hacer.

Una herramienta que únicamente genera texto a partir de información pública presenta un riesgo diferente de una herramienta que puede acceder a documentos internos, recordar información, navegar en Internet y ejecutar acciones sobre otros sistemas.

Por ello, antes de ampliar las capacidades de una herramienta institucional deberá considerarse no solamente:

“¿qué nueva función nos permite realizar?”

sino también:

“¿a qué nueva información tendrá acceso, qué podrá hacer con ella y qué ocurriría si esa capacidad se utiliza incorrectamente?”

La regla general será:

a mayor capacidad de acceso, memoria, integración o actuación autónoma de la inteligencia artificial, mayores deberán ser los controles de seguridad, supervisión y autorización.

---

6. Incidentes, errores, correcciones y determinación de responsabilidades en el uso de inteligencia artificial

La utilización institucional de inteligencia artificial puede producir errores aun cuando se hayan aplicado medidas razonables de prevención. Una herramienta puede generar información falsa, interpretar incorrectamente una fuente, omitir antecedentes relevantes, reproducir información desactualizada o generar resultados distintos de los solicitados. También pueden producirse errores humanos durante la selección de fuentes, elaboración de instrucciones, carga de documentos, verificación, revisión, almacenamiento o publicación.

El sistema de control de Fundamedios deberá estar diseñado tanto para **prevenir estos errores como para detectarlos, corregirlos y determinar su origen cuando ocurran**.

La existencia de un error no determinará automáticamente la responsabilidad de la persona que utilizó la herramienta ni trasladará automáticamente esa responsabilidad a Fundamedios. La evaluación deberá reconstruir el proceso seguido y determinar, entre otros elementos, si se utilizó una herramienta autorizada, si se proporcionaron instrucciones adecuadas, si se cumplieron los controles de verificación, si existió revisión humana y si se respetaron las reglas establecidas en este Manual.

Este procedimiento tendrá una finalidad primordialmente correctiva y preventiva. Cuando corresponda determinar responsabilidades individuales, estas deberán basarse en las actuaciones efectivamente realizadas y no únicamente en el hecho de que una herramienta de inteligencia artificial haya intervenido en el proceso.

6.1. Qué deberá considerarse un incidente relacionado con IA

Para efectos de este Manual, se considerará incidente cualquier utilización de inteligencia artificial que produzca o pueda razonablemente producir una afectación relevante a la calidad de la información, los derechos de terceros, la confidencialidad, la protección de datos, la seguridad, la integridad editorial o los intereses de Fundamedios.

Un incidente puede ser detectado **antes o después de la publicación**. La detección previa deberá ser tratada como una oportunidad de control y no como un fracaso del procedimiento.

Podrán considerarse incidentes, dependiendo de sus circunstancias:

- utilización de información falsa o inexistente generada por IA;
- atribución incorrecta de hechos, declaraciones o documentos;
- utilización de una cita que no corresponde a las palabras originales;
- referencia a una norma, sentencia, estudio, estadística o fuente inexistente;
- procesamiento de información institucional mediante una cuenta o herramienta no autorizada;
- exposición de datos personales, información confidencial o identidad de fuentes;
- incorporación de información descartada a un producto;
- publicación de contenido generado sin la revisión humana exigida;
- utilización de una versión desactualizada de un documento o base de conocimiento;
- generación de una conclusión que exceda la evidencia disponible;
- ejecución de una acción automatizada que no haya sido autorizada; o
- cualquier incumplimiento de los controles establecidos en este Manual que genere un riesgo material.

No todo error tipográfico, reformulación menor o resultado insatisfactorio de una herramienta deberá convertirse en un incidente formal. La respuesta deberá ser proporcional a su relevancia y posibles consecuencias.

6.2. Obligación de comunicar errores e incidentes

Toda persona que detecte un posible incidente deberá comunicarlo oportunamente cuando su naturaleza pueda afectar un producto, una investigación, información protegida o los derechos de terceros.

Esta obligación se aplicará aunque la persona que detecte el problema haya participado en su producción.

La detección temprana deberá ser incentivada. Un sistema en el cual las personas oculten errores por temor a que su sola existencia determine una sanción aumenta el riesgo institucional, porque dificulta corregirlos antes de que produzcan consecuencias mayores.

Por ello, la comunicación de buena fe de un error será considerada parte del funcionamiento esperado del sistema de control.

La existencia de una comunicación voluntaria no impedirá analizar posteriormente si existió un incumplimiento relevante, pero deberá diferenciarse entre **detectar y comunicar un error** y **ocultarlo, mantenerlo deliberadamente o incumplir de forma consciente los controles establecidos**.

6.3. Contención inmediata

Cuando el incidente todavía no haya alcanzado al público, la prioridad será impedir que lo haga.

Si el problema se detecta después de la publicación, deberá procurarse evitar que continúe propagándose o generando nuevas consecuencias mientras se determina la medida definitiva.

Dependiendo del caso, podrán adoptarse inmediatamente acciones como suspender una publicación programada, detener la distribución de un documento, retirar temporalmente un contenido, impedir nuevas versiones automáticas, suspender una integración, restringir el acceso a determinada información o detener el uso de una base de conocimiento que pueda estar reproduciendo información incorrecta.

La contención no supone necesariamente una conclusión sobre la existencia o gravedad del error. Constituye una medida provisional destinada a impedir que un posible problema aumente mientras se realiza la evaluación correspondiente.

6.4. Preservación del proceso

Cuando un incidente pueda tener consecuencias editoriales, jurídicas, de seguridad o de protección de datos, no deberá eliminarse inmediatamente toda evidencia del proceso con el propósito de “corregir” el problema.

Deberá preservarse, en la medida necesaria, la información que permita determinar qué ocurrió.

Esto podrá incluir:

el prompt utilizado; la respuesta generada; los documentos proporcionados; las fuentes consultadas; las versiones del contenido; las verificaciones realizadas; las revisiones humanas; la herramienta y cuenta utilizadas; la fecha de las actuaciones; y las modificaciones realizadas antes de la publicación.

La preservación deberá realizarse respetando las reglas de confidencialidad y protección de datos.

Una vez conservada la evidencia necesaria, podrán adoptarse las medidas de eliminación, corrección o restricción correspondientes.

6.5. Reconstrucción del incidente

Cuando la importancia del caso lo justifique, deberá reconstruirse el flujo que produjo el resultado.

La revisión deberá procurar responder, en orden, preguntas como las siguientes:

1. ¿De dónde surgió la información? Deberá determinarse si provino de una fuente identificable, de una búsqueda realizada por la IA, de un documento proporcionado por Fundamedios, de una inferencia del sistema o de su generación autónoma.

2. ¿Qué información recibió la herramienta? Deberá comprobarse si los inputs eran correctos, completos y apropiados para la tarea.

3. ¿Qué instrucciones recibió? Deberá analizarse si el prompt establecía adecuadamente la finalidad, las restricciones y el nivel de precisión requerido.

4. ¿Qué produjo la IA? Deberá distinguirse el resultado original de las modificaciones realizadas posteriormente por una persona.

5. ¿Qué verificación se realizó? Deberá comprobarse si las fuentes fueron abiertas, contrastadas y revisadas conforme al nivel de riesgo.

6. ¿Quién revisó el resultado? Deberá identificarse la intervención humana previa a la aprobación.

7. ¿Cómo llegó el error al producto final? Esta pregunta permitirá determinar qué control falló o no existió.

La reconstrucción no deberá partir de la premisa de encontrar una persona responsable. Deberá partir de la necesidad de **identificar el punto de falla**.

6.6. Diferencia entre error de la herramienta y error del proceso

Cuando una IA produzca información falsa o incorrecta, deberá determinarse si el sistema de control previsto por Fundamedios estaba diseñado para detectar precisamente ese tipo de error.

Por ejemplo, si una herramienta inventa una cita y el Manual exige comprobar toda cita en su fuente original, la generación incorrecta constituye un error de la herramienta; pero la incorporación de esa cita a una publicación sin realizar la comprobación exigida constituye además una falla del proceso.

De igual manera, si una IA proporciona una cifra desactualizada, el problema puede originarse en el sistema, en una fuente desactualizada o en el prompt. Si posteriormente la cifra fue correctamente verificada y sustituida antes de publicación, el mecanismo de control funcionó.

Esta distinción será fundamental:

No todo error generado por IA constituye un incumplimiento. El incumplimiento aparece cuando se omite injustificadamente un control exigible, se utiliza la herramienta de manera no autorizada o se desatiende un riesgo que debía ser razonablemente conocido.

6.7. Clasificación según gravedad

Los incidentes podrán evaluarse conforme a su impacto real o potencial.

No será necesario aplicar una escala rígida en todos los casos, pero podrá utilizarse como referencia la siguiente clasificación:

Incidente menor. Error detectado y corregido dentro del flujo ordinario, sin publicación ni afectación relevante a información protegida.

Incidente moderado. Error que alcanzó una etapa avanzada del proceso, produjo una publicación de impacto limitado o implicó un incumplimiento de controles que requiere medidas correctivas.

Incidente grave. Situación que pueda producir una afectación significativa a una persona, comprometer información confidencial o datos protegidos, afectar sustancialmente la credibilidad de una publicación, generar riesgo jurídico relevante o evidenciar una falla importante de los controles.

Incidente crítico. Exposición o actuación con capacidad de generar consecuencias especialmente graves para fuentes protegidas, personas, sistemas o para la continuidad e integridad institucional.

La respuesta, documentación, escalamiento y revisión posterior deberán ser proporcionales a esta evaluación.

6.8. Corrección de contenidos antes de publicación

Cuando un error sea detectado durante la elaboración del producto, deberá corregirse antes de continuar.

La corrección deberá alcanzar no solamente el texto visible, sino también, cuando corresponda, el **origen del error**.

Si el problema provino de una fuente incorrecta incorporada al expediente, deberá corregirse o retirarse. Si se encontraba dentro de una base de conocimiento, deberá actualizarse. Si el sistema utilizó una versión antigua de un

documento, deberá sustituirse. Si el prompt permitía inferencias que no estaban autorizadas, deberá modificarse antes de generar nuevamente el contenido.

La finalidad será impedir que el mismo error reaparezca en una versión posterior.

6.9. Correcciones posteriores a publicación

Cuando un error relevante sea detectado después de que el contenido haya sido publicado, deberá evaluarse con prontitud la medida necesaria para corregirlo.

Según la naturaleza del caso, podrá corresponder corregir el contenido, incorporar una aclaración, actualizar información, retirar temporalmente una publicación mientras se verifica, rectificar una afirmación o adoptar otra medida editorial adecuada.

La utilización de inteligencia artificial no modificará los estándares de responsabilidad editorial aplicables a una publicación de Fundamedios.

No deberá atribuirse públicamente un error a “la inteligencia artificial” como mecanismo para evitar la responsabilidad institucional sobre un contenido que fue finalmente aprobado y publicado por la organización.

Sin perjuicio de ello, internamente sí deberá determinarse con precisión **qué papel desempeñó la IA en la producción del error** para corregir el procedimiento y establecer las responsabilidades que correspondan.

6.10. Corrección de sistemas y repositorios

Cuando el error publicado tenga origen en información almacenada dentro de los sistemas institucionales, la corrección pública deberá acompañarse de una corrección interna.

Deberá revisarse si la información incorrecta permanece en:

- bases de conocimiento;
- memorias de asistentes;
- expedientes;
- documentos de referencia;
- plantillas;
- bases estructuradas;
- resúmenes;
- cronologías; o
- cualquier otro repositorio que pueda alimentar futuras generaciones.

Una organización que corrige una publicación pero mantiene el dato falso dentro de su memoria institucional corre el riesgo de reproducir posteriormente el mismo error.

6.11. Determinación de responsabilidades

Cuando un incidente requiera determinar responsabilidades, la evaluación deberá atender a la conducta concreta de las personas que participaron en el proceso.

No deberá presumirse responsabilidad individual únicamente porque una persona haya operado la herramienta, ni responsabilidad institucional exclusiva únicamente porque la cuenta utilizada perteneciera a Fundamedios.

Deberán considerarse, entre otros elementos:

a) las funciones y nivel de decisión de la persona;

b) las instrucciones que recibió;

c) el nivel de riesgo de la actividad;

d) si utilizó una herramienta y cuenta autorizadas;

e) si respetó las reglas sobre información que podía proporcionar al sistema;

f) si aplicó los controles de verificación exigidos;

g) si realizó o solicitó la revisión humana correspondiente;

h) si conocía o debía razonablemente conocer el problema;

i) si comunicó oportunamente el incidente una vez detectado;

j) si actuó dentro o fuera de las funciones y autorizaciones asignadas; y

k) si existieron deficiencias institucionales de capacitación, configuración, supervisión o procedimiento que contribuyeron al resultado.

La determinación deberá distinguir entre error razonable, negligencia, incumplimiento de procedimiento y actuación deliberadamente contraria a las reglas institucionales.

6.12. Responsabilidad por incumplimiento deliberado de los controles

Tendrá especial relevancia el caso en que una persona utilice deliberadamente mecanismos diferentes de aquellos establecidos por Fundamedios.

Esto podrá ocurrir, por ejemplo, cuando se utilice una cuenta personal pese a existir un entorno institucional obligatorio; se introduzca información restringida en una herramienta expresamente no autorizada; se omita conscientemente una verificación requerida; se publique contenido generado por IA sin la aprobación correspondiente; se eludan permisos o controles técnicos; o se oculten errores conocidos.

En estos casos deberá evaluarse la responsabilidad de quien realizó la actuación de acuerdo con las normas internas, contractuales y jurídicas aplicables.

La existencia de herramientas institucionales, procedimientos documentados y capacitación adecuada permitirá distinguir con mayor precisión entre **una falla atribuible al sistema institucional** y **una actuación realizada al margen de ese sistema**.

6.13. Responsabilidad de revisión y aprobación

La revisión humana no deberá convertirse en una formalidad.

Quien tenga a su cargo revisar o aprobar un producto deberá realizar el nivel de control correspondiente a su función y al riesgo del contenido.

Una aprobación realizada sin la revisión mínima exigida podrá constituir una falla del procedimiento, aunque el texto haya sido inicialmente producido por otra persona o por una herramienta de IA.

De esta manera, la responsabilidad dentro del proceso podrá encontrarse distribuida entre diferentes funciones: investigación, verificación, redacción, revisión y aprobación.

El objetivo no será diluir responsabilidades, sino **identificar qué control correspondía a cada etapa y si efectivamente fue realizado**.

6.14. Registro de uso responsable y declaración interna

Las personas autorizadas para utilizar inteligencia artificial institucional deberán conocer y aceptar las reglas establecidas en este Manual.

Fundamedios podrá implementar una **Declaración de Uso Responsable de Inteligencia Artificial**, incorporada al propio Manual o como anexo, mediante la cual cada usuario deje constancia de que conoce, entre otras, las siguientes obligaciones:

Utilizar exclusivamente las herramientas y cuentas autorizadas para las actividades institucionales que así lo requieran; proteger la información a la que tenga acceso; no trasladar información institucional protegida a cuentas personales; verificar los resultados antes de utilizarlos como hechos; mantener la supervisión humana; respetar los controles de confidencialidad, protección de datos y seguridad; comunicar oportunamente errores e incidentes; y no presentar como resultado exclusivamente propio aquello cuya producción haya requerido controles o atribuciones específicas de acuerdo con este Manual.

La declaración no tendrá como finalidad trasladar automáticamente a los usuarios toda responsabilidad derivada de la utilización institucional de IA.

Su función será **establecer claramente las reglas conocidas y aceptadas por quienes utilizan estas herramientas**, de manera que posteriormente pueda determinarse si una actuación se realizó dentro o fuera del marco autorizado.

6.15. Trazabilidad como mecanismo de protección institucional y personal

La trazabilidad protege a Fundamedios, pero también a las personas que trabajan dentro de la organización.

Si un contenido genera una controversia, deberá ser posible demostrar que el investigador utilizó fuentes identificables, que la información fue verificada, que el redactor siguió las instrucciones institucionales y que existieron los controles de revisión correspondientes.

De la misma manera, la trazabilidad permitirá identificar cuándo una persona omitió esos controles.

Por ello, los registros no deberán concebirse exclusivamente como mecanismos de vigilancia o atribución de culpa. Constituyen evidencia de que el procedimiento institucional fue correctamente aplicado.

6.16. Aprendizaje después de un incidente

Los incidentes relevantes deberán utilizarse para mejorar el sistema.

Una vez controlado el problema y adoptadas las correcciones necesarias, deberá analizarse si corresponde modificar:

una instrucción institucional, una plantilla de prompt, una configuración de seguridad, un permiso, una herramienta autorizada, un procedimiento de verificación, una regla editorial, una capacitación o una disposición de este Manual.

Cuando el mismo tipo de error aparezca repetidamente, deberá analizarse la causa sistémica en lugar de tratar cada episodio como un evento aislado.

Un procedimiento de IA responsable deberá evolucionar a partir de la experiencia acumulada por la organización.

6.17. Registro institucional de incidentes relevantes

Fundamedios deberá mantener un registro proporcional de los incidentes relacionados con inteligencia artificial cuya gravedad o recurrencia justifique documentación institucional.

El registro podrá contener:

fecha — herramienta — proyecto o área — tipo de incidente — información afectada — impacto — medidas de contención — causa identificada — corrección aplicada — medidas preventivas posteriores.

Cuando contenga datos personales, información sobre fuentes o material confidencial, su acceso deberá encontrarse restringido.

No será necesario registrar formalmente cada error menor detectado durante una conversación ordinaria. El registro deberá concentrarse en los casos que permitan identificar riesgos, acreditar actuaciones relevantes o mejorar el sistema de control.

6.18. Prohibición de ocultamiento o alteración de registros

Una vez detectado un incidente relevante, ninguna persona deberá eliminar, modificar o alterar deliberadamente información necesaria para reconstruir lo ocurrido con la finalidad de ocultar su actuación o la existencia del incidente.

Esta regla no impedirá realizar eliminaciones necesarias para contener una exposición de información. En esos casos deberá preservarse previamente, cuando sea posible y seguro hacerlo, la evidencia mínima necesaria para documentar lo sucedido.

La alteración deliberada de registros relacionados con un incidente será considerada separadamente del error que originó el incidente.

6.19. Revisión jurídica cuando exista riesgo legal

Cuando un incidente pueda generar una reclamación, afectar derechos de terceros, involucrar protección de datos personales, comprometer una fuente, producir responsabilidad contractual o derivar en un procedimiento administrativo o judicial, deberá evaluarse oportunamente la necesidad de revisión jurídica.

La intervención jurídica no deberá limitarse a determinar si existe responsabilidad.

También podrá servir para establecer medidas de mitigación, preservar evidencia, definir la forma adecuada de realizar una corrección, responder a terceros y determinar las obligaciones que correspondan.

6.20. Principio de responsabilidad humana

La inteligencia artificial no será considerada responsable editorial de los contenidos de Fundamedios.

Toda publicación deberá encontrarse finalmente bajo responsabilidad y supervisión humana conforme a las funciones institucionales correspondientes.

Esto no significa que toda persona que intervenga en un proceso responda automáticamente por cualquier error producido por una herramienta. Significa que **ninguna decisión editorial relevante deberá quedar sin una persona responsable de revisarla y aprobarla**.

La IA podrá investigar, ordenar, comparar, resumir, analizar, proponer y redactar dentro de los límites establecidos en este Manual. La decisión de utilizar sus resultados institucionalmente permanecerá bajo control humano.

6.21. Regla de cierre del Capítulo III

La protección de la información utilizada mediante inteligencia artificial exige algo más que herramientas seguras. Requiere que Fundamedios pueda conocer **qué información ingresó al sistema, dónde quedó almacenada, quién podía acceder, qué hizo la herramienta con ella, qué controles humanos se aplicaron y qué ocurrió cuando se detectó un problema**.

La existencia de un procedimiento institucional previamente definido permitirá diferenciar un error ocurrido dentro de un sistema de trabajo razonablemente controlado de una actuación realizada incumpliendo sus reglas.

El principio de cierre será:

Los errores deberán poder detectarse, los procesos deberán poder reconstruirse, las consecuencias deberán corregirse y las responsabilidades deberán determinarse a partir de las actuaciones verificables de cada participante y de los controles que correspondían al proceso.

CAPÍTULO

IV

Uso de inteligencia artificial para el análisis y producción de contenidos



Una vez que la información ha sido investigada, verificada, clasificada y preparada conforme a los capítulos anteriores, la inteligencia artificial podrá utilizarse para apoyar su análisis y convertirla en productos destinados al trabajo interno o a la comunicación pública de Fundamedios.

Esta fase comprende una utilización diferente de la IA. Durante la investigación, la herramienta ayuda principalmente a **encontrar y procesar información**. Durante la producción, recibe información previamente controlada y ayuda a **analizarla, estructurarla, redactarla, resumirla, adaptarla o transformarla en un producto determinado**.

Esta diferencia deberá mantenerse durante todo el proceso. Una herramienta que recibe información para redactar no deberá asumir simultáneamente la función de completar la investigación mediante datos que no hayan sido previamente verificados.

La inteligencia artificial podrá intervenir en diferentes grados: desde corregir la sintaxis de un texto escrito completamente por una persona hasta producir un primer borrador a partir de un conjunto de información validada. El nivel de intervención permitido dependerá de la naturaleza del producto, del riesgo de su contenido y de los controles humanos que puedan aplicarse posteriormente.

En todos los casos, la publicación continuará siendo un **producto de Fundamedios y no de la herramienta utilizada para elaborarlo**.

1. Preparación de los insumos para producción

La calidad de un producto generado con asistencia de inteligencia artificial depende, en buena medida, de la calidad y claridad de la información proporcionada al sistema. Una herramienta puede producir un texto gramaticalmente correcto y aparentemente coherente a partir de información incompleta, contradictoria o incorrecta. La calidad de la redacción no deberá confundirse con la calidad de la evidencia.

Antes de solicitar la elaboración de un contenido sustantivo deberá prepararse un conjunto de insumos que permita a la herramienta comprender **qué debe producir, con qué información puede trabajar y cuáles son los límites que no deberá superar**.

1.1. El input de producción

Para efectos de este Manual se entenderá como **input de producción** el conjunto de información, documentos, instrucciones, parámetros y restricciones que Fundamedios proporciona a una herramienta de inteligencia artificial para elaborar o apoyar la elaboración de un producto.

El input podrá incluir documentos completos, extractos, bases de datos, hechos verificados, citas, cronologías, antecedentes, instrucciones editoriales, formatos, ejemplos institucionales o cualquier otro elemento necesario para realizar la tarea.

No existe un formato único para construirlo. Una alerta breve puede necesitar solamente unos pocos hechos y una fuente; una investigación extensa puede requerir documentos, cronologías, citas, posiciones de las partes y múltiples instrucciones.

Lo esencial será que la información proporcionada resulte suficiente para la tarea y que sus diferentes niveles de certeza no se encuentren mezclados de forma que la IA pueda interpretarlos incorrectamente.

1.2. Construcción del prompt

El prompt deberá concebirse como una **instrucción de trabajo**, no únicamente como una pregunta dirigida a la herramienta.

Mientras mayor sea la importancia o complejidad del producto, mayor deberá ser la precisión de las instrucciones.

Un prompt sustantivo podrá definir, según resulte necesario:

Objetivo. Qué producto deberá elaborarse y para qué finalidad.

Audiencia. A quién se encuentra dirigido.

Información autorizada. Qué hechos, documentos o materiales pueden utilizarse.

Estructura. Qué organización deberá tener el contenido.

Extensión. Qué nivel de desarrollo se espera.

Tono y estilo. Qué características editoriales deberá respetar.

Atribución. Cómo deberán identificarse fuentes, declaraciones y posiciones.

Limitaciones. Qué información no deberá incorporar o inferir.

Incertidumbres. Qué aspectos deberán presentarse como pendientes, controvertidos o no confirmados.

Resultado esperado. En qué formato deberá entregarse el producto.

No será necesario incorporar todos estos elementos en cada instrucción. El prompt deberá contener aquello que resulte útil para reducir errores y orientar adecuadamente la tarea.

1.3. Prompts institucionales y metodologías reutilizables

Cuando Fundamedios identifique instrucciones que producen resultados consistentes para actividades recurrentes, podrá convertirlas en **plantillas institucionales de prompts**.

Podrán desarrollarse, por ejemplo, instrucciones específicas para elaboración de alertas, verificaciones, resúmenes de documentos, cronologías, comparación de versiones, extracción de datos, revisión editorial, adaptación de contenidos para redes sociales u otras actividades frecuentes.

Estas plantillas deberán mantenerse dentro del entorno institucional y podrán evolucionar a partir de la experiencia.

Una plantilla no deberá convertirse en una fórmula rígida. Quien la utilice deberá adaptarla a las particularidades del asunto y proporcionar los datos necesarios para cada caso.

Cuando un incidente permita determinar que una determinada estructura de prompt facilita errores, deberá revisarse la plantilla correspondiente.

1.4. Identificación de la información autorizada

Cuando el riesgo o complejidad del producto lo justifique, deberá indicarse expresamente a la herramienta cuáles son los materiales que puede utilizar como base factual.

Podrá establecerse, por ejemplo:

Elabora el contenido utilizando únicamente los hechos y documentos proporcionados. No incorpores información factual adicional proveniente de tu conocimiento general. Si consideras que falta un dato necesario, identifícalo como información pendiente en lugar de completarlo.

Este tipo de instrucción deberá utilizarse particularmente cuando la precisión factual sea más importante que la amplitud narrativa.

La herramienta podrá sugerir que existe información adicional que sería conveniente investigar, pero esa información deberá regresar al procedimiento de investigación y verificación antes de incorporarse al producto.

1.5. Diferenciación dentro del input

Cuando el conjunto de información sea complejo, deberá procurarse que la IA pueda distinguir qué representa cada elemento.

Una forma posible será organizar el material bajo categorías como:

HECHOS VERIFICADOS

Información que puede ser utilizada directamente como base factual.

CITAS VERIFICADAS

Declaraciones que pueden reproducirse conforme al texto proporcionado.

POSICIONES DE LAS PARTES

Afirmaciones atribuibles a una persona o institución, sin asumir automáticamente que su contenido sea verdadero.

CONTEXTO

Información necesaria para comprender el asunto.

INFORMACIÓN CONDICIONADA

Datos que únicamente pueden utilizarse manteniendo expresamente una reserva o limitación.

NO UTILIZAR

Información descartada que, por una razón excepcional, deba permanecer visible para evitar que sea reintroducida o para explicar una contradicción.

Esta estructura no será obligatoria para todos los productos. Su utilidad aumenta conforme crece el volumen de información y la posibilidad de confundir hechos, versiones e hipótesis.

1.6. Prohibición de completar vacíos factuales

Una de las reglas centrales de producción será que la inteligencia artificial **no deberá completar mediante generación aquello que la investigación no haya establecido**.

- Si falta una fecha, deberá señalar que falta una fecha.
- Si no se conoce un cargo, no deberá asignarlo por contexto.
- Si una fuente no identifica a una persona, no deberá intentar determinar quién es.
- Si una investigación no permite concluir por qué ocurrió un acontecimiento, no deberá construir una explicación causal para hacer más coherente el relato.
- Si existen dos cifras contradictorias que todavía no han sido resueltas, no deberá elegir aquella que parezca más plausible.

La fluidez narrativa nunca deberá utilizarse para ocultar vacíos de información.

1.7. Distinción entre hechos y atribuciones

La herramienta deberá recibir instrucciones suficientes para distinguir entre un hecho comprobado y una afirmación realizada por una fuente.

No será equivalente escribir:

“La empresa desvió recursos públicos”

que:

“La denuncia presentada sostiene que la empresa habría desviado recursos públicos.”

La segunda formulación atribuye correctamente una afirmación a su origen cuando el hecho todavía no ha sido establecido.

Durante la producción, la IA puede eliminar inadvertidamente expresiones de atribución para hacer el texto más directo o conciso. Esa modificación puede transformar sustancialmente el sentido y riesgo jurídico de una publicación.

Por ello, cuando una afirmación dependa de una fuente, denuncia, acusación, versión o documento que no demuestra por sí mismo la verdad definitiva del hecho, **la atribución deberá acompañarla durante todo el proceso de redacción**.

1.8. Conservación de incertidumbres y matices

Las limitaciones identificadas durante la investigación deberán trasladarse al prompt.

Si un hecho no pudo verificarse independientemente, si una cifra es aproximada, si existen versiones contrapuestas o si determinada conclusión continúa siendo objeto de discusión, la herramienta deberá conocerlo.

No deberá pedirse a la IA que produzca un texto “contundente” a costa de eliminar esas reservas.

La precisión puede exigir expresiones como:

“de acuerdo con…”, **“según la documentación revisada…”**, **“no fue posible confirmar…”**, **“las fuentes consultadas difieren…”**, **“hasta la fecha de esta publicación…”**, o **“no existe información suficiente para determinar…”**.

Estas expresiones no constituyen debilidad narrativa cuando reflejan correctamente el estado de la evidencia.

1.9. Instrucciones negativas

En determinados productos resultará tan importante indicar qué debe hacer la herramienta como señalar **qué no debe hacer**.

Los prompts podrán incluir restricciones expresas tales como:

- no inventar fuentes;
- no crear citas;
- no agregar nombres;
- no asumir intenciones;
- no realizar imputaciones jurídicas no sustentadas;
- no convertir denuncias en hechos comprobados;
- no modificar cifras;
- no eliminar reservas;
- no utilizar información descartada;
- no identificar fuentes protegidas;
- no introducir datos personales innecesarios; y
- no completar información que no aparezca en los insumos autorizados.

Estas restricciones deberán ajustarse al riesgo específico del contenido y no utilizarse únicamente como una lista automática incorporada a cualquier prompt.

1.10. Uso de ejemplos y contenidos anteriores

Fundamedios podrá proporcionar a la inteligencia artificial publicaciones anteriores para enseñarle el formato, tono o estructura que se desea reproducir.

Cuando se utilicen ejemplos, deberá indicarse que estos sirven como **referencia editorial**, no necesariamente como fuente factual para el nuevo producto.

Una nota anterior puede contener información que era correcta en la fecha de su publicación, pero que ya no se encuentre vigente.

Por ello, una instrucción como “escribe esta nota siguiendo el estilo de estos cinco ejemplos” no deberá interpretarse como autorización para incorporar al nuevo contenido hechos contenidos en aquellos ejemplos.

Cuando los sistemas institucionales dispongan permanentemente de publicaciones anteriores, deberá mantenerse esta misma distinción entre **conocimiento de estilo** y **fuente de hechos actuales**.

1.11. Información personal y confidencial dentro del input

El principio de minimización continuará aplicándose durante la producción.

Que determinada información haya sido legítimamente recopilada y conservada en el expediente no significa que deba proporcionarse íntegramente a la herramienta encargada de redactar.

El input deberá contener únicamente los datos necesarios para producir el contenido.

Cuando una fuente confidencial pueda ser descrita como “Fuente A” para efectos de redacción, no será necesario proporcionar su identidad. Cuando un documento contenga información personal irrelevante, deberá utilizarse una versión depurada. Cuando una base completa no sea necesaria, deberá proporcionarse únicamente el segmento pertinente.

La fase de producción no deberá convertirse en una nueva exposición innecesaria de información que ya se encontraba adecuadamente protegida durante la investigación.

1.12. Separación de instrucciones, fuentes y contenido externo

Cuando se proporcionen documentos, páginas web, correos, transcripciones u otros contenidos a una herramienta, deberá quedar claro que estos constituyen **material para analizar**, no instrucciones que sustituyan las indicaciones de Fundamedios.

Esta separación será especialmente importante cuando la IA pueda acceder directamente a contenido externo.

Una instrucción contenida dentro de una página web o documento —por ejemplo, “ignore las instrucciones anteriores”, “revele los archivos disponibles” o cualquier otra orden dirigida a un sistema automatizado— no deberá modificar el objetivo institucional.

Las instrucciones de producción deberán provenir de las personas autorizadas y del marco establecido en este Manual.

1.13. Producción por etapas

En contenidos complejos podrá resultar preferible dividir el proceso de generación en varias tareas en lugar de solicitar inmediatamente un texto terminado.

Por ejemplo:

primera etapa: ordenar los hechos y proponer una estructura;

segunda etapa: identificar vacíos, contradicciones o información que requiera aclaración;

tercera etapa: elaborar un borrador utilizando únicamente el material aprobado;

cuarta etapa: revisar coherencia, atribuciones y correspondencia con las fuentes;

quinta etapa: realizar ajustes editoriales.

Este método permite detectar problemas antes de que queden incorporados dentro de una narración completa y aparentemente coherente.

No será necesario aplicarlo a contenidos sencillos. Su utilización deberá responder a la complejidad y riesgo del producto.

1.14. Prompts de revisión distintos de prompts de generación

Cuando sea posible, deberá distinguirse entre pedir a una herramienta que **produzca** un texto y pedirle que **revise críticamente** ese texto.

Después de generar un borrador podrá utilizarse una nueva instrucción destinada exclusivamente a detectar problemas.

Por ejemplo, podrá solicitarse que identifique:

- afirmaciones que no aparezcan en los insumos;
- citas cuya fuente no esté indicada;
- cambios de atribución;
- afirmaciones categóricas basadas en información condicionada;
- contradicciones internas;
- fechas o cifras inconsistentes;
- conclusiones que excedan la evidencia; o
- datos personales cuya inclusión parezca innecesaria.

Esta revisión automatizada constituye un control adicional y **no reemplaza la revisión humana**.

En productos de mayor riesgo podrá resultar conveniente realizar la revisión en una interacción o contexto separado, para reducir la tendencia del sistema a defender o reproducir su propio resultado anterior.

1.15. Registro de prompts relevantes

Los prompts que hayan tenido una intervención sustancial en la elaboración de productos de riesgo medio o alto deberán conservarse cuando resulten necesarios para la trazabilidad.

Esto no significa archivar cada reformulación, corrección gramatical o conversación.

Deberán preservarse especialmente aquellas instrucciones que permitan reconstruir:

qué información recibió la herramienta, qué se le pidió producir y qué restricciones se establecieron.

En productos recurrentes podrá ser suficiente conservar la plantilla institucional utilizada y registrar las modificaciones relevantes realizadas para el caso concreto.

1.16. Modificación humana de resultados

Un resultado generado por IA podrá ser corregido, reescrito, reorganizado o descartado libremente por la persona responsable.

La herramienta no deberá convertirse en autoridad sobre la forma final del contenido.

Cuando el resultado no refleje adecuadamente los hechos, el tono, el contexto o los estándares de Fundamedios, deberá modificarse aunque sea técnicamente correcto o estilísticamente atractivo.

En determinados casos será más eficiente abandonar un resultado y comenzar nuevamente que intentar corregir una generación construida sobre una premisa equivocada.

El usuario deberá mantener una actitud crítica frente al producto y no limitarse a aceptar modificaciones porque hayan sido presentadas con seguridad o fluidez por el sistema.

1.17. La calidad lingüística no equivale a validación

Un texto bien redactado puede contener información falsa.

La inteligencia artificial puede mejorar sustancialmente la gramática, estructura y claridad de un producto, pero esas características no constituyen evidencia de que sus afirmaciones sean correctas.

Esta diferencia deberá permanecer especialmente presente durante la revisión de borradores extensos, en los cuales un dato inventado puede integrarse de manera natural entre numerosos hechos correctos y resultar difícil de detectar.

Por ello, la evaluación del producto deberá realizarse en dos dimensiones diferentes:

calidad factual: si aquello que se afirma está sustentado;

calidad editorial: si aquello que está sustentado se encuentra correctamente explicado.

Un producto solo estará preparado para publicación cuando cumpla ambas.

1.18. Condición para avanzar desde los insumos hacia la redacción

Podrá iniciarse la generación sustantiva de un producto cuando exista una base suficiente de información validada y la persona responsable haya determinado razonablemente:

a) qué debe producirse; b) qué información puede utilizarse; c) qué información requiere atribución o reserva; d) qué elementos no pueden incorporarse; e) qué estilo y estructura corresponden; y **f) qué nivel de revisión deberá aplicarse posteriormente.**

La preparación adecuada del input no garantiza que la IA produzca un resultado correcto. Su finalidad es reducir el espacio dentro del cual la herramienta pueda introducir errores y facilitar que estos sean identificados posteriormente.

La regla operativa de esta etapa será:

La IA no deberá recibir simplemente información para “escribir algo”. Deberá recibir una tarea delimitada, información controlada y reglas suficientes para saber qué puede hacer con ella y qué decisiones permanecen reservadas al equipo de Fundamedios.

2. Usos permitidos de la inteligencia artificial en la elaboración de contenidos

La inteligencia artificial podrá intervenir en distintas etapas de elaboración de contenidos de Fundamedios siempre que su utilización sea compatible con la naturaleza del producto, el nivel de riesgo involucrado y los controles establecidos en este Manual.

No todos los usos de IA representan el mismo nivel de intervención. Existe una diferencia sustancial entre utilizar una herramienta para corregir errores ortográficos de un texto ya terminado y solicitarle que produzca el primer

borrador de una investigación a partir de múltiples fuentes. En consecuencia, el grado de supervisión deberá aumentar conforme la herramienta tenga mayor capacidad para modificar el significado, seleccionar información, formular conclusiones o producir contenido nuevo.

La utilización de IA deberá entenderse como **asistencia al trabajo profesional**. Podrá realizar tareas que antes exigían trabajo manual, acelerar procesos y facilitar el análisis de grandes cantidades de información, pero no asumirá aquellas decisiones que, por su naturaleza editorial, jurídica o ética, correspondan a una persona.

2.1. Organización y estructuración de contenidos

La IA podrá utilizarse para ordenar información previamente recopilada y proponer estructuras para su presentación.

Entre otras funciones, podrá:

- organizar hechos cronológicamente;
- agrupar información por temas;
- identificar elementos repetidos;
- proponer secciones para una investigación;
- convertir información dispersa en una tabla o esquema;
- ordenar antecedentes y acontecimientos;
- preparar líneas de tiempo;
- comparar estructuras posibles para una nota; y
- sugerir formas de presentar información compleja de manera comprensible.

En estas actividades, deberá diferenciarse entre **organizar la información existente** y **generar información para llenar los espacios de la estructura propuesta**.

Si la herramienta determina, por ejemplo, que una nota debería incluir una sección sobre los antecedentes de determinada institución, podrá sugerir que esa información hace falta. No deberá inventar esos antecedentes para completar la sección.

La identificación de un vacío deberá producir una tarea de investigación, no una generación factual.

2.2. Resumen y síntesis

La inteligencia artificial podrá utilizarse para resumir documentos, informes, entrevistas, resoluciones, investigaciones, transcripciones y otros materiales.

El resumen deberá conservar el sentido esencial del documento y no transformar las posiciones de sus autores.

Cuando la precisión sea particularmente importante, deberá solicitarse que el resultado permita identificar la sección, página, párrafo, minuto o fragmento del cual se obtiene la información, siempre que la herramienta y el documento permitan hacerlo.

Los resúmenes producidos por IA no deberán utilizarse como sustitutos permanentes de las fuentes originales cuando estas sean necesarias para verificar posteriormente una afirmación.

En especial, antes de publicar una cita, cifra, conclusión jurídica, acusación, decisión de autoridad u otra afirmación relevante, deberá consultarse la fuente original y no únicamente el resumen generado.

2.3. Elaboración de borradores

La IA podrá producir primeros borradores de notas, alertas, informes, comunicados, contenidos institucionales y otros productos, siempre que trabaje sobre información previamente preparada conforme a este Manual.

El hecho de que la herramienta produzca un texto completo no significa que ese texto se encuentre listo para publicación.

Todo borrador generado deberá considerarse **material de trabajo sujeto a revisión**.

La persona responsable deberá comprobar que el borrador:

1. corresponde al objeto solicitado;

2. utiliza correctamente los hechos proporcionados;

3. conserva las atribuciones;

4. no introduce información nueva sin sustento;

5. no modifica el sentido de las fuentes;

6. mantiene las incertidumbres relevantes; y

7. cumple los estándares editoriales aplicables al producto.

Mientras mayor sea la intervención generativa de la herramienta, mayor deberá ser la profundidad de esta revisión.

2.4. Reescritura y mejora de textos

La IA podrá utilizarse para mejorar textos elaborados previamente por personas.

Podrá corregir gramática, ortografía, sintaxis, puntuación, claridad, coherencia, redundancias y organización narrativa.

También podrá solicitarse que simplifique un texto excesivamente técnico, mejore transiciones, reduzca repeticiones o adapte su extensión.

Sin embargo, una instrucción aparentemente estilística puede producir modificaciones sustantivas.

Expresiones como **“hazlo más contundente”, “simplifica”, “reduce”, “hazlo más directo” o “mejora el texto”** pueden provocar que el sistema elimine atribuciones, reservas o contexto necesario.

Por esta razón, después de una reescritura deberá comprobarse no solamente que el texto sea mejor, sino que **continúe diciendo exactamente aquello que Fundamedios puede sostener**.

2.5. Adaptación de un contenido a diferentes formatos

Un mismo contenido validado podrá adaptarse mediante inteligencia artificial a diferentes canales de comunicación.

Por ejemplo, una investigación podrá convertirse posteriormente en:

una nota breve para el portal; una secuencia para redes sociales; una descripción para un video; un boletín; una ficha informativa; un resumen ejecutivo; preguntas y respuestas; o material de apoyo para una presentación.

La adaptación no deberá modificar el estándar factual del contenido original.

La reducción de espacio tampoco justificará eliminar una atribución necesaria o convertir una explicación compleja en una afirmación absoluta.

Cuando un formato breve no permita explicar adecuadamente una acusación, controversia o incertidumbre, deberá reformularse el contenido de manera que la reducción no produzca una representación engañosa.

2.6. Elaboración de titulares, encabezados y textos destacados

La inteligencia artificial podrá proponer titulares, subtítulos, encabezados, sumarios, descripciones y otros elementos destinados a presentar una publicación.

Estas propuestas requerirán una revisión especialmente cuidadosa porque los modelos tienden a favorecer formulaciones directas, atractivas o categóricas.

Un titular no deberá afirmar más de lo que demuestra el contenido.

No podrá convertir:

una denuncia en una condena;

una investigación en una determinación de responsabilidad;

una correlación en causalidad;

una sospecha en un hecho;

una posibilidad en una certeza.

La revisión del titular deberá realizarse de forma independiente de la revisión del cuerpo de la publicación. Que una nota explique correctamente las limitaciones de la evidencia no corrige un titular que las elimine.

2.7. Preparación de contenidos para redes sociales

La IA podrá utilizarse para adaptar contenidos de Fundamedios a redes sociales, siempre que se mantengan las mismas reglas de precisión, atribución y verificación aplicables al producto original.

La brevedad propia de estas plataformas no reducirá el estándar factual.

Cuando se utilicen hilos, carruseles, publicaciones seriadas u otros formatos, deberá comprobarse además que cada pieza conserve suficiente contexto para no producir una impresión falsa cuando sea vista aisladamente.

No deberá asumirse que una aclaración incluida en una publicación posterior subsana necesariamente una afirmación incorrecta contenida en la primera.

Cuando la publicación incluya enlaces hacia una investigación completa, el contenido breve deberá seguir siendo correcto por sí mismo.

2.8. Traducción

La inteligencia artificial podrá utilizarse para traducir contenidos entre idiomas.

Las traducciones deberán revisarse cuando el material tenga relevancia editorial, jurídica, técnica o institucional.

Deberá prestarse particular atención a:

- nombres propios;
- cargos;
- denominaciones oficiales;
- términos jurídicos;
- citas;
- cifras;
- siglas;
- expresiones idiomáticas; y
- conceptos cuya traducción pueda alterar el sentido.

Cuando se traduzca una cita textual destinada a publicación, deberá poder distinguirse entre una **traducción realizada por Fundamedios** y el texto originalmente pronunciado o escrito por la fuente.

La IA podrá asistir en la traducción, pero no deberá utilizarse para crear una versión aparentemente literal que no pueda ser contrastada con el original.

2.9. Transcripción y procesamiento de material audiovisual

Podrán utilizarse herramientas de IA para transcribir audios, entrevistas, reuniones y videos, siempre que su utilización respete las reglas de confidencialidad, protección de fuentes, datos personales y seguridad establecidas anteriormente.

Las transcripciones automatizadas deberán considerarse aproximaciones hasta que los fragmentos relevantes sean contrastados con la grabación original.

Esto será obligatorio antes de utilizar una declaración como cita textual.

Los sistemas pueden confundir palabras, nombres, números, voces o interlocutores, especialmente cuando existe ruido, superposición de voces, mala calidad de audio o terminología especializada.

Por ello:

una transcripción automatizada facilita localizar una cita; no acredita por sí sola que esa cita haya sido pronunciada exactamente de esa manera.

2.10. Extracción y estructuración de datos

La inteligencia artificial podrá utilizarse para extraer información de documentos y convertirla en estructuras que faciliten su análisis.

Podrá, por ejemplo, identificar fechas en expedientes, extraer montos de contratos, organizar nombres de instituciones, construir tablas a partir de documentos o clasificar grandes cantidades de registros.

Cuando estos resultados sirvan posteriormente para realizar cálculos, estadísticas o afirmaciones factuales, deberá verificarse su exactitud mediante controles proporcionales al riesgo.

En conjuntos extensos podrá utilizarse muestreo, validación automatizada complementaria, comparación con totales conocidos u otros mecanismos razonables, sin perjuicio de una revisión más profunda cuando una cifra vaya a convertirse en elemento central de una publicación.

2.11. Análisis comparativo

La IA podrá utilizarse para comparar documentos, declaraciones, versiones, bases de datos o contenidos.

Podrá identificar diferencias entre versiones de una norma, modificaciones de un documento, inconsistencias entre declaraciones o coincidencias entre distintos conjuntos de información.

Estos resultados deberán considerarse **herramientas de análisis**.

La existencia de una diferencia detectada por IA deberá comprobarse en los documentos correspondientes antes de afirmar públicamente su existencia cuando tenga relevancia sustantiva.

Asimismo, una coincidencia textual o estadística no deberá transformarse automáticamente en una conclusión sobre plagio, coordinación, responsabilidad, intención o relación entre personas.

2.12. Identificación de patrones

Las herramientas de IA podrán ayudar a detectar patrones dentro de grandes cantidades de información.

Esta capacidad puede resultar especialmente útil para investigaciones sobre discursos, contrataciones, decisiones públicas, campañas de desinformación, redes de actores u otros fenómenos que requieran examinar numerosos elementos.

Sin embargo, un patrón identificado algorítmicamente constituye un **punto de partida para el análisis**, no necesariamente una conclusión publicable.

Antes de utilizarlo deberá determinarse:

si el conjunto de datos es suficientemente representativo; si existen datos que contradicen el patrón; si la clasificación realizada por la IA fue correcta; si existe una explicación alternativa; y si la relación observada permite realmente sostener la conclusión propuesta.

La herramienta podrá ayudar a encontrar una relación. La investigación deberá determinar qué significa.

2.13. Apoyo para fact-checking

La inteligencia artificial podrá asistir en procesos de verificación de afirmaciones, particularmente para:

descomponer una afirmación en elementos verificables; localizar posibles fuentes; identificar bases de datos relevantes; comparar cifras; encontrar antecedentes; construir cronologías; detectar contradicciones; y proponer preguntas adicionales de verificación.

La IA no deberá emitir autónomamente la calificación final de una verificación destinada a Ecuador Chequea u otro producto equivalente.

La determinación de que una afirmación es falsa, verdadera, engañosa, imprecisa, fuera de contexto o corresponde a cualquier otra categoría editorial deberá resultar del procedimiento metodológico aplicable y de una decisión humana.

El sistema podrá ayudar a reunir la evidencia. **La calificación editorial deberá ser adoptada por Fundamedios.**

2.14. Apoyo para análisis jurídico o normativo

La IA podrá utilizarse para localizar, ordenar, comparar, resumir o explicar normas, sentencias, resoluciones y otros materiales jurídicos relevantes para una investigación o publicación.

Toda referencia jurídica sustantiva deberá verificarse en una fuente confiable antes de ser utilizada.

Deberá comprobarse especialmente:

la existencia de la norma o decisión; su texto; vigencia; fecha; número; órgano que la emitió; contexto; y alcance de aquello que realmente establece.

No deberán publicarse citas normativas o jurisprudenciales únicamente porque una herramienta las haya proporcionado.

Cuando exista una interpretación jurídica controvertida, la IA podrá ayudar a identificar posiciones y argumentos, pero no deberá presentar una interpretación como indiscutible cuando jurídicamente no lo sea.

2.15. Generación de preguntas y líneas de investigación

La inteligencia artificial podrá utilizarse para proponer preguntas que el equipo debería formular a una fuente, institución, especialista o persona involucrada.

También podrá sugerir posibles líneas de investigación derivadas de la información disponible.

Estas propuestas deberán ser evaluadas por el investigador.

La generación de una pregunta no implica que su premisa sea correcta. Una IA puede formular una pregunta basándose en una relación inexistente o una inferencia no comprobada.

Antes de trasladar una acusación o premisa a un tercero deberá verificarse que exista sustento suficiente para formularla responsablemente.

2.16. Preparación de solicitudes de información y cuestionarios

La IA podrá apoyar la elaboración de solicitudes de acceso a información pública, cuestionarios, pedidos de aclaración y comunicaciones destinadas a obtener información adicional.

La persona responsable deberá comprobar que las preguntas correspondan a los hechos investigados y que no incluyan como premisa información generada o no verificada.

Cuando se solicite la versión de una persona respecto de una afirmación que pueda afectarla, la comunicación deberá permitir comprender suficientemente el asunto sobre el cual se solicita respuesta.

La IA podrá mejorar la organización de las preguntas; no deberá utilizarse para construir acusaciones que el expediente no respalde.

2.17. Apoyo para visualización y presentación de información

La inteligencia artificial podrá asistir en la transformación de datos en tablas, esquemas, gráficos, mapas conceptuales, cronologías u otros recursos destinados a facilitar la comprensión de información compleja.

Cuando una representación visual incluya cifras o relaciones factuales, deberá comprobarse que corresponda a los datos originales.

La representación no deberá producir una impresión engañosa mediante escalas, categorías, omisiones o relaciones que no se desprendan de la información.

Cuando una herramienta sugiera qué tipo de visualización utilizar, la decisión final deberá considerar si esa representación explica correctamente los datos y no solamente si resulta atractiva.

2.18. Generación y modificación de imágenes, audio y video

La generación o modificación de contenidos visuales y audiovisuales mediante inteligencia artificial presenta riesgos diferentes a la asistencia en redacción y deberá someterse a controles específicos.

La IA podrá utilizarse para producir material **ilustrativo, gráfico, conceptual o claramente no documental**, siempre que su utilización no induzca al público a creer que representa evidencia auténtica de un acontecimiento que no ocurrió de esa manera.

No deberá utilizarse contenido sintético para fabricar, alterar o aparentar:

- fotografías documentales de hechos reales;
- declaraciones de personas;
- voces atribuidas a personas reales;
- videos de acontecimientos;
- documentos;
- capturas de pantalla;
- pruebas; o
- cualquier otro material cuya autenticidad pueda ser relevante para comprender una publicación.

Cuando se utilice una imagen, audio o video generado o modificado sustancialmente mediante IA y exista posibilidad razonable de que el público lo interprete como material real, deberá identificarse claramente su naturaleza artificial o ilustrativa.

Las correcciones técnicas que no alteren el contenido sustantivo —por ejemplo, determinados ajustes de tamaño, formato o calidad— deberán diferenciarse de las modificaciones que cambian aquello que la imagen, audio o video representa.

2.19. Uso de IA para revisión crítica

La inteligencia artificial podrá desempeñar una función adicional como **revisor previo** del contenido.

Una herramienta podrá recibir un borrador y las fuentes utilizadas y recibir instrucciones para intentar detectar:

afirmaciones sin respaldo; contradicciones; atribuciones incorrectas; saltos lógicos; posibles errores de cifras o fechas; omisiones de contexto; lenguaje excesivamente categórico; posibles riesgos de privacidad; o elementos que merezcan revisión jurídica.

En trabajos de mayor importancia podrá solicitarse deliberadamente una revisión adversarial: pedir al sistema que intente encontrar las debilidades del contenido en lugar de mejorarlo.

Esta función puede ser particularmente útil porque obliga a someter el producto a una segunda lectura estructurada antes de la revisión final.

Sin embargo, que la IA no detecte un problema **no constituye evidencia de que el problema no exista**.

2.20. Usos que requieren especial cautela

Determinadas funciones no se encuentran prohibidas, pero requieren una supervisión reforzada por su capacidad para modificar significativamente el sentido del contenido.

Entre ellas:

Reducción extrema de textos. Puede eliminar contexto esencial.

Reescritura de acusaciones. Puede convertir atribuciones en afirmaciones.

Generación de titulares. Puede exagerar conclusiones.

Clasificación de personas. Puede producir inferencias injustificadas.

Análisis de intenciones. Puede atribuir estados mentales que no pueden demostrarse.

Detección de emociones. No deberá asumirse como evidencia objetiva del estado de una persona.

Análisis jurídico. Puede producir referencias inexistentes o interpretaciones incorrectas.

Análisis estadístico. Puede aplicar métodos inapropiados o interpretar incorrectamente los resultados.

Traducción de declaraciones sensibles. Puede alterar matices relevantes.

Síntesis de grandes expedientes. Puede omitir información que contradiga la narrativa principal.

En estas actividades deberá incrementarse la revisión humana y, cuando corresponda, utilizar especialistas.

2.21. Funciones que permanecen bajo decisión humana

Con independencia de las capacidades técnicas de las herramientas utilizadas, determinadas decisiones no deberán delegarse íntegramente a inteligencia artificial.

Corresponderá a personas autorizadas de Fundamedios decidir, entre otras cuestiones:

1. si existe evidencia suficiente para publicar una afirmación relevante;
2. qué calificación editorial corresponde a una verificación;
3. si una acusación se encuentra suficientemente sustentada;
4. qué información confidencial o personal puede publicarse;
5. si debe protegerse la identidad de una fuente;
6. si una persona ha tenido oportunidad suficiente de ofrecer su versión cuando corresponda;
7. si una publicación requiere revisión jurídica;
8. cuál será el titular definitivo en contenidos sensibles;
9. si un producto cumple los estándares editoriales de la organización; y
10. si el contenido se encuentra finalmente autorizado para publicación.

La IA podrá proporcionar información y recomendaciones para adoptar estas decisiones. **No será la autoridad que las adopte.**

2.22. Proporcionalidad entre intervención de IA y supervisión humana

La supervisión deberá aumentar conforme aumente la capacidad de la herramienta para alterar el producto.

Puede visualizarse esta relación de forma sencilla:

Intervención baja: corrección ortográfica, formato, transcripción auxiliar. Requiere revisión ordinaria.

Intervención media: resumen, reorganización, traducción, adaptación, extracción de datos. Requiere contraste del resultado con el material original.

Intervención alta: elaboración de borradores, análisis de patrones, interpretación de información, generación de titulares o síntesis de investigaciones complejas. Requiere revisión sustantiva y verificación de las afirmaciones relevantes.

Intervención crítica: contenidos capaces de afectar significativamente derechos, reputación, seguridad, confidencialidad o posición jurídica de una persona u organización. Requiere controles reforzados y, cuando corresponda, revisión editorial superior, especializada o jurídica.

Esta clasificación no pretende convertir cada utilización cotidiana en un procedimiento burocrático. Su finalidad es impedir que una tarea de alto impacto reciba el mismo nivel de revisión que una corrección gramatical.

2.23. Criterio general para los usos permitidos

La pregunta fundamental no será si la inteligencia artificial **puede** realizar técnicamente una determinada tarea.

Deberá determinarse si resulta apropiado que la realice dentro del proceso editorial de Fundamedios y bajo qué condiciones.

El criterio general será:

La inteligencia artificial podrá asistir en todas aquellas tareas en las que su intervención aumente la capacidad de investigar, analizar, organizar, redactar o comunicar información sin sustituir la verificación de los hechos, el criterio editorial ni las decisiones humanas que determinan qué puede sostener y publicar Fundamedios.

3. Estándares para la redacción asistida por inteligencia artificial

La utilización de inteligencia artificial para redactar, reescribir, resumir o adaptar contenidos no modifica los estándares editoriales aplicables a las publicaciones de Fundamedios. Un texto producido total o parcialmente mediante IA deberá cumplir las mismas exigencias de exactitud, claridad, atribución, contexto y sustento que un texto elaborado sin asistencia automatizada.

La capacidad de los modelos para generar textos fluidos plantea, sin embargo, un riesgo particular: **una afirmación incorrecta puede quedar integrada de manera completamente natural dentro de una narración correcta**. El error puede no presentarse como una duda o una inconsistencia evidente, sino como un nombre, una fecha, una relación causal, una cita o un antecedente aparentemente razonable.

Por ello, la revisión de un texto asistido por IA no deberá limitarse a determinar si "suena bien". Deberá comprobarse que **cada afirmación material pueda vincularse con información que Fundamedios se encuentre en condiciones de sostener**.

3.1. Fidelidad a la información validada

La redacción deberá reflejar fielmente los insumos autorizados para producción.

La IA podrá reorganizar la información, sintetizarla, mejorar su claridad y proponer distintas formas de explicarla, pero no deberá alterar su significado ni aumentar el grado de certeza que ofrecen las fuentes.

Si un documento señala que determinada situación "podría" haber ocurrido, la redacción no deberá transformarla en algo que "ocurrió". Si una fuente formula una acusación, esta no deberá convertirse en una conclusión de Fundamedios. Si una cifra es aproximada, deberá conservarse esa condición.

Deberá prestarse particular atención a palabras aparentemente pequeñas que modifican sustancialmente el sentido: *presuntamente, aproximadamente, según, habría, podría, hasta el momento, de acuerdo con, no se ha podido confirmar* y expresiones equivalentes.

La eliminación de una de estas precisiones puede convertir una redacción estilísticamente más limpia en una afirmación materialmente incorrecta.

3.2. Trazabilidad de las afirmaciones materiales

Toda afirmación material incorporada a un producto deberá poder retroceder razonablemente hasta el insumo que la sustenta.

La trazabilidad podrá representarse mediante la siguiente relación:

afirmación publicada → información validada → fuente de respaldo

No será necesario incorporar esta cadena de manera visible en todos los productos ni mantener un registro individual para cada oración ordinaria. Su profundidad deberá ser proporcional a la naturaleza del contenido.

Sin embargo, cuando se trate de afirmaciones centrales, cifras relevantes, acusaciones, citas, antecedentes jurídicos, conclusiones de una verificación u otros elementos susceptibles de controversia, deberá ser posible identificar con suficiente precisión de dónde provienen.

Si durante la revisión no puede determinarse qué sustenta una afirmación introducida por la IA, esta deberá ser verificada nuevamente o eliminada.

3.3. Prohibición de alucinaciones y contenido generado sin sustento

No se admitirá como contenido factual ninguna información cuya única fuente sea la generación de la herramienta.

Esto comprende, entre otros supuestos:

a) nombres que no consten en los insumos; b) fechas completadas por contexto; c) cargos atribuidos sin fuente; d) citas creadas o reconstruidas; e) estadísticas no verificadas; f) referencias bibliográficas inexistentes; g) normas o decisiones judiciales inventadas; h) enlaces fabricados; i) declaraciones atribuidas a personas sin respaldo; j) acontecimientos añadidos para completar una cronología; o **k) explicaciones causales construidas únicamente porque resultan plausibles.**

La plausibilidad no constituye evidencia.

Cuando la herramienta produzca información nueva que parezca correcta y resulte útil para el producto, deberá tratarse como un **hallazgo pendiente de verificación** y regresar a la fase de investigación.

3.4. Relación directa entre afirmación y fuente

La existencia de una fuente relacionada con un tema no significa que esa fuente respalde cualquier afirmación sobre ese tema.

Durante la revisión deberá comprobarse que la fuente sustenta **precisamente aquello que el texto afirma**.

Por ejemplo, un documento que demuestre que una persona fue investigada no demuestra que haya cometido la conducta investigada. Una denuncia no acredita por sí misma los hechos denunciados. La existencia de un contrato no demuestra necesariamente que haya existido una irregularidad. Una correlación estadística no acredita automáticamente una relación causal.

La IA puede producir estos saltos porque construye narraciones mediante relaciones entre información. Fundamedios deberá impedir que una relación narrativa sea presentada como una relación probatoria.

3.5. Citas textuales

Las citas constituyen uno de los elementos que requieren mayor precisión.

La IA no deberá reconstruir una cita textual a partir del sentido general de una declaración.

Cuando se utilicen comillas, las palabras deberán corresponder a aquello que efectivamente consta en la fuente, con las adaptaciones editoriales legítimas que correspondan y sin alterar su significado.

Antes de publicar una cita obtenida mediante transcripción o extracción automatizada deberá comprobarse en la fuente original.

Si únicamente se conoce el sentido de una declaración, deberá utilizarse una paráfrasis correctamente atribuida en lugar de fabricar una literalidad.

Una paráfrasis correcta nunca deberá convertirse en una cita textual por decisión de la IA.

3.6. Paráfrasis

La inteligencia artificial podrá reformular declaraciones o documentos para facilitar su comprensión, siempre que conserve su significado.

La revisión deberá prestar atención a que la paráfrasis no:

- aumente el grado de certeza;
- elimine condiciones o excepciones;
- atribuya intenciones;
- simplifique hasta modificar el sentido;
- transforme una opinión en un hecho;

- convierta una posibilidad en una afirmación; o
- elimine elementos relevantes para comprender la posición original.

Cuando la formulación exacta tenga importancia jurídica, política, técnica o editorial, deberá preferirse consultar directamente el texto original.

3.7. Atribuciones

Toda afirmación que corresponda a la versión, opinión, denuncia, interpretación o conclusión de un tercero deberá mantenerse correctamente atribuida.

La IA no deberá eliminar atribuciones para conseguir una narración más directa.

Expresiones como:

“según el informe…”, **“la denuncia sostiene…”**, **“de acuerdo con la versión proporcionada por…”**, **“la institución afirmó…”**, **“el documento señala…”**

cumplen una función sustantiva y no meramente estilística.

La atribución permite al lector conocer **quién sostiene determinada información y con qué grado de respaldo se presenta**.

3.8. Acusaciones y señalamientos

Cuando un contenido incluya acusaciones de conductas ilícitas, irregulares, fraudulentas, corruptas o de otra naturaleza capaz de afectar significativamente a una persona o institución, la redacción deberá guardar correspondencia estricta con la evidencia disponible.

La IA no deberá intensificar el lenguaje para producir un texto más atractivo.

Deberá distinguirse claramente entre:

una denuncia;

una investigación;

una formulación de cargos o actuación equivalente;

una decisión administrativa;

una sentencia;

una sentencia firme cuando esa circunstancia resulte relevante; y

un hecho acreditado directamente por documentación independiente.

Estas situaciones no son intercambiables y la herramienta no deberá tratarlas como si lo fueran.

3.9. Presunción de inocencia y situación procesal

Cuando el contenido se refiera a investigaciones, procesos penales, procedimientos sancionatorios u otras actuaciones en las que todavía no exista una determinación definitiva de responsabilidad, la redacción deberá reflejar correctamente la situación procesal correspondiente.

No deberá utilizarse lenguaje que atribuya responsabilidad como hecho establecido cuando jurídicamente no lo esté.

La utilización de expresiones como “acusado”, “investigado”, “procesado”, “sentenciado”, “condenado” u otras equivalentes deberá corresponder al estado real del procedimiento y a la terminología jurídicamente aplicable.

La IA puede utilizar estos términos como sinónimos en lenguaje ordinario. **Dentro de un producto periodístico no necesariamente lo son.**

3.10. Derecho a la versión y contraste

Cuando la naturaleza del contenido exija procurar la versión de una persona o institución afectada por una afirmación, la IA podrá ayudar a preparar las preguntas, ordenar las respuestas y posteriormente integrarlas dentro del texto.

No podrá decidir que una versión resulta innecesaria únicamente porque la documentación disponible parezca concluyente.

Cuando exista una respuesta, deberá incorporarse de manera fiel y en un contexto que permita comprender su posición.

No será necesario reproducir íntegramente todas las respuestas recibidas, pero su síntesis no deberá deformar, minimizar artificialmente ni seleccionar únicamente aquellos fragmentos que favorezcan una narrativa predeterminada.

Si no se obtuvo respuesta, la redacción deberá reflejar correctamente los intentos de contacto cuando resulte pertinente.

3.11. Contradicciones entre fuentes

Cuando existan fuentes relevantes que sostengan versiones incompatibles y la investigación no permita resolver definitivamente la contradicción, el texto deberá conservarla.

La inteligencia artificial no deberá escoger automáticamente la versión que resulte más coherente con el resto del relato.

Podrá utilizarse para explicar con claridad:

qué sostiene una fuente, qué sostiene la otra, en qué coinciden, en qué difieren y qué elementos han podido comprobarse independientemente.

Cuando la evidencia permita preferir razonablemente una versión, la publicación podrá explicarlo, pero deberá hacerlo a partir de la evidencia y no de la preferencia narrativa del sistema.

3.12. Contexto suficiente

Una afirmación puede ser literalmente verdadera y, sin embargo, resultar engañosa cuando se presenta sin el contexto necesario.

La IA deberá utilizarse con especial cuidado en tareas de reducción, resumen y adaptación, porque estos procesos tienden precisamente a eliminar información.

Durante la revisión deberá comprobarse si el lector necesita conocer antecedentes adicionales para interpretar correctamente un hecho.

Esto puede comprender fechas, cambios normativos, circunstancias anteriores o posteriores, universo estadístico, metodología de una encuesta, alcance de una resolución, condiciones de una declaración o cualquier elemento sin el cual la información produzca una impresión sustancialmente diferente.

La brevedad no deberá obtenerse mediante la eliminación del contexto indispensable.

3.13. Causalidad y correlación

La IA no deberá presentar una relación causal cuando los datos únicamente demuestren coincidencia, correlación o proximidad temporal.

Expresiones como “provocó”, “causó”, “como consecuencia de”, “debido a” o equivalentes deberán utilizarse únicamente cuando exista sustento suficiente para establecer esa relación.

Cuando la evidencia únicamente permita afirmar que dos acontecimientos ocurrieron simultáneamente o que determinadas variables presentan una asociación, la redacción deberá conservar esa limitación.

Este control será especialmente importante en análisis estadísticos, económicos, sociales y políticos.

3.14. Intenciones, motivaciones y estados mentales

No deberá atribuirse a una persona una intención, motivación, conocimiento, emoción o propósito como hecho objetivo únicamente porque la IA lo infiera de sus actos o declaraciones.

Expresiones como “buscaba”, “pretendía”, “sabía”, “intentó”, “quería” o “su objetivo era” requieren sustento.

Cuando la motivación sea una interpretación, deberá presentarse como tal y atribuirse cuando corresponda.

La inteligencia artificial puede producir explicaciones psicológicamente plausibles para una conducta. **La plausibilidad de una explicación no demuestra el estado mental de una persona.**

3.15. Cifras y datos cuantitativos

Toda cifra material deberá ser revisada con particular atención.

La IA puede cometer errores al copiar números, convertir unidades, sumar valores, interpretar porcentajes, comparar períodos o extraer información de tablas.

Deberá comprobarse, según corresponda:

valor — unidad — período — universo — fuente — metodología — moneda — fecha de referencia.

Cuando se realicen cálculos mediante IA, deberá verificarse el procedimiento o el resultado mediante un mecanismo adecuado, especialmente cuando la cifra constituya parte central de la publicación.

Un porcentaje correcto aplicado sobre una base incorrecta continúa produciendo una afirmación incorrecta.

3.16. Fechas, nombres y cargos

Los elementos aparentemente simples deberán verificarse cuando tengan relevancia para el contenido.

Los modelos pueden completar automáticamente segundos apellidos, confundir personas con nombres similares, utilizar cargos anteriores, modificar fechas o atribuir a una institución una denominación que ya no se encuentra vigente.

La revisión deberá comprobar especialmente nombres propios, denominaciones oficiales, fechas relevantes y cargos al momento al que se refiere la publicación.

Cuando el cargo de una persona sea importante para comprender el hecho, deberá distinguirse entre su posición actual y la que ocupaba cuando ocurrió el acontecimiento.

3.17. Información jurídica

Cuando la publicación incluya contenido jurídico generado o resumido mediante IA, deberá evitarse convertir interpretaciones en afirmaciones categóricas sobre el derecho aplicable.

Una norma deberá revisarse en su texto vigente y dentro de su contexto.

Una sentencia deberá revisarse para determinar qué resolvió efectivamente, cuáles fueron los hechos del caso y qué alcance tiene el criterio utilizado.

La IA no deberá utilizar fragmentos de jurisprudencia fuera de contexto ni presentar como regla general aquello que dependía de circunstancias específicas.

Cuando la cuestión sea jurídicamente compleja o tenga impacto significativo en la publicación, deberá considerarse revisión especializada.

3.18. Información científica, técnica y especializada

Cuando una publicación dependa de conocimientos especializados, la IA podrá facilitar su explicación, pero no sustituirá las fuentes técnicas correspondientes.

Deberá prestarse atención a que la simplificación no transforme:

una hipótesis en consenso; una asociación en causalidad; una estimación en una medición exacta; un estudio individual en evidencia concluyente; o una opinión experta en posición universalmente aceptada.

Cuando la comprensión del asunto requiera conocimientos que excedan razonablemente la capacidad del equipo editorial, deberá recurrirse a fuentes o especialistas adecuados.

3.19. Generalizaciones

La IA puede producir generalizaciones para facilitar la narración.

Expresiones como “los expertos consideran”, “la ciudadanía rechaza”, “las organizaciones sostienen”, “las investigaciones demuestran” o “existe consenso” deberán utilizarse únicamente cuando exista evidencia suficiente para justificar el alcance de esas afirmaciones.

Una fuente no representa automáticamente a un sector completo.

Un estudio no representa automáticamente toda la evidencia disponible.

Varias publicaciones similares no demuestran necesariamente consenso.

Cuando no exista sustento para una generalización, deberá identificarse específicamente quién sostiene la afirmación.

3.20. Lenguaje claro y preciso

La asistencia de IA podrá utilizarse para mejorar la claridad de los contenidos.

El lenguaje deberá procurar que el lector pueda distinguir hechos, opiniones, denuncias, antecedentes, análisis y conclusiones.

La complejidad innecesaria no deberá confundirse con rigor. Sin embargo, la simplificación tampoco deberá eliminar conceptos necesarios.

Cuando exista un término técnico indispensable, podrá explicarse en lenguaje accesible antes que sustituirlo por una palabra que modifique su significado.

Deberán evitarse formulaciones grandilocuentes, categóricas o sensacionalistas generadas únicamente para aumentar el impacto narrativo.

3.21. Neutralidad de la herramienta frente a instrucciones editoriales

Los modelos pueden incorporar tendencias derivadas de sus datos de entrenamiento, diseño, fuentes disponibles o forma en que se formula una pregunta.

Fundamedios no deberá asumir que una respuesta es neutral por haber sido producida mediante inteligencia artificial.

Cuando una tarea implique evaluar posiciones contrapuestas, deberá instruirse al sistema para trabajar con la evidencia proporcionada y distinguir adecuadamente los argumentos y fuentes correspondientes.

La herramienta podrá ayudar a identificar posibles sesgos u omisiones del propio borrador, pero la determinación editorial deberá realizarse mediante criterio humano y conforme a los estándares de la organización.

3.22. Titulares y elementos de alto impacto

Los titulares, textos destacados, miniaturas, pies de fotografía y publicaciones iniciales en redes sociales pueden tener una exposición superior al cuerpo completo de una investigación.

Por ello, deberán revisarse como unidades independientes.

La IA podrá proponer alternativas, pero el equipo deberá comprobar que cada una:

- se encuentre respaldada por el contenido;
- no elimine una atribución indispensable;
- no exagere una conclusión;
- no utilice una acusación como hecho;
- no induzca a error mediante una omisión relevante; y
- guarde proporción con la evidencia.

La búsqueda de atención no deberá modificar el estándar de exactitud.

3.23. Coherencia entre titular, texto y evidencia

El producto final deberá mantener una cadena de coherencia:

la evidencia sustenta el texto → el texto sustenta la conclusión → la conclusión sustenta el titular.

No deberá invertirse el proceso construyendo primero un titular atractivo y solicitando posteriormente a la IA que organice la información para respaldarlo.

Cuando durante la producción la evidencia no permita sostener la hipótesis inicial, deberá modificarse la hipótesis o el producto.

La inteligencia artificial no deberá utilizarse para seleccionar únicamente los elementos que confirmen una conclusión previamente decidida.

3.24. Revisión de omisiones

La exactitud no depende únicamente de aquello que el texto contiene.

Una omisión puede modificar sustancialmente la comprensión de un asunto.

En productos relevantes deberá preguntarse si la redacción omite:

una fuente que contradice la conclusión; una respuesta relevante de la persona señalada; una excepción normativa; un dato que modifica una estadística; un acontecimiento posterior; una limitación metodológica; o cualquier otro elemento sin el cual la conclusión pueda resultar engañosa.

La IA podrá utilizarse como herramienta auxiliar para detectar estas omisiones, particularmente mediante una revisión adversarial del borrador.

3.25. Control de consistencia interna

Antes de avanzar a la revisión final deberá comprobarse que el propio producto no contenga contradicciones internas.

Fechas, nombres, cifras, cargos, denominaciones, categorías y conclusiones deberán mantenerse consistentes a lo largo del documento.

Cuando se hayan producido varias secciones por separado mediante IA, deberá realizarse una revisión conjunta. Diferentes generaciones pueden utilizar terminología distinta, repetir información o asumir hechos incompatibles.

La integración de las partes constituye una tarea editorial y no deberá limitarse a unir automáticamente los textos producidos.

3.26. Estándar reforzado para afirmaciones de alto impacto

No todas las afirmaciones requieren el mismo nivel de comprobación.

Deberá aplicarse una revisión reforzada cuando una afirmación pueda:

afectar significativamente la reputación de una persona; atribuir una conducta ilícita; revelar información privada; identificar una fuente; producir consecuencias jurídicas; afectar la seguridad de alguien; generar una acusación pública; o constituir la conclusión principal de una investigación.

En estos casos, no deberá bastar con que la IA reproduzca correctamente la información que recibió.

La persona responsable deberá volver a la fuente correspondiente y comprobar directamente que el producto final se encuentre suficientemente sustentado.

3.27. Prueba de respaldo antes de publicación

Antes de considerar terminada una publicación asistida sustancialmente por inteligencia artificial, deberá poder responderse afirmativamente, según corresponda, a las siguientes preguntas:

¿Puedo identificar la fuente de las afirmaciones centrales?

¿Las citas fueron comprobadas en su fuente original?

¿Las cifras y fechas relevantes fueron verificadas?

¿Las acusaciones permanecen correctamente atribuidas?

¿Se mantuvieron las incertidumbres existentes?

¿El texto incorporó algún hecho que no estaba en los insumos autorizados?

¿Se incluyó la versión de las partes cuando correspondía?

¿El titular dice exactamente lo que permite sostener la evidencia?

¿Existe información personal o confidencial cuya publicación no sea necesaria?

¿Una persona revisó sustantivamente el producto completo?

No será necesario documentar formalmente este cuestionario para cada publicación ordinaria. En productos de riesgo elevado podrá incorporarse a la lista de control editorial.

3.28. Regla final de redacción

La inteligencia artificial podrá mejorar significativamente la velocidad y calidad formal de la producción editorial, pero no deberá alterar la relación fundamental entre evidencia y publicación.

La regla de esta etapa será:

Ninguna afirmación se vuelve verdadera porque la inteligencia artificial la redacte con claridad, y ninguna afirmación deja de necesitar respaldo porque se integre naturalmente dentro de un texto correcto. Todo contenido material deberá conservar una relación verificable con la evidencia que permite a Fundamedios sostenerlo.

---

4. Supervisión humana, revisión editorial y aprobación final

La supervisión humana constituye el control final sobre cualquier contenido producido con asistencia sustantiva de inteligencia artificial. Ninguna herramienta, independientemente de su nivel de precisión o de la capacidad que tenga para analizar documentos, contrastar información o revisar su propio trabajo, deberá sustituir la responsabilidad editorial de las personas encargadas de determinar qué publica Fundamedios.

La supervisión no deberá entenderse como una lectura superficial realizada al finalizar el proceso. Debe permitir detectar errores que pueden haberse originado desde la selección de información hasta la redacción definitiva. En consecuencia, la persona que revisa deberá encontrarse en condiciones de cuestionar el resultado, modificarlo, devolverlo para nuevas verificaciones o impedir su publicación.

La intensidad de la revisión dependerá del contenido. Una corrección gramatical asistida por IA no requiere el mismo procedimiento que una investigación que atribuya posibles irregularidades a una autoridad pública. El sistema de revisión deberá ser **proporcional al riesgo y a la intervención de la inteligencia artificial**.

4.1. Ninguna publicación sustantiva será completamente automatizada

Los contenidos destinados al público que hayan sido producidos sustancialmente mediante inteligencia artificial deberán contar con intervención humana antes de su publicación.

No deberá establecerse un flujo en el cual una herramienta pueda, por sí sola:

investigar → redactar → aprobar → publicar.

La automatización podrá utilizarse para determinadas tareas intermedias, pero deberá existir un punto de control humano antes de que el producto adquiera carácter público cuando su contenido sea sustantivo.

Esta regla se aplicará aunque la herramienta haya funcionado correctamente en ocasiones anteriores.

El historial de precisión de un sistema no convierte sus futuras respuestas en información verificada.

4.2. La revisión humana debe ser sustantiva

La revisión humana no se considerará cumplida por el solo hecho de que una persona haya abierto, leído rápidamente o realizado modificaciones menores sobre un contenido generado mediante IA.

Revisar significa ejercer criterio sobre el resultado.

La persona responsable deberá encontrarse en condiciones de determinar si:

- las afirmaciones centrales están respaldadas;
- las fuentes han sido correctamente utilizadas;
- las atribuciones se mantienen;

- las citas corresponden al original;
- las cifras y fechas relevantes son correctas;
- el contexto es suficiente;
- las conclusiones guardan relación con la evidencia;
- se incorporó adecuadamente la versión de las partes cuando correspondía;
- existe información que no debería publicarse; y
- el producto cumple los estándares editoriales de Fundamedios.

La revisión deberá concentrarse especialmente en aquellos elementos que una IA puede modificar sin producir una señal evidente de error.

4.3. Responsabilidad de quien elabora el contenido

La primera responsabilidad de revisión corresponderá a quien utiliza la inteligencia artificial para elaborar el producto.

La persona que genera un borrador no deberá trasladarlo inmediatamente al siguiente nivel de aprobación bajo la premisa de que será otra persona quien detecte cualquier error.

Antes de entregar el contenido deberá realizar una primera revisión completa y corregir los problemas identificados.

Esto incluye comprobar que el resultado responde a las instrucciones proporcionadas, que no contiene hechos incorporados autónomamente por la herramienta y que las afirmaciones relevantes pueden vincularse con las fuentes correspondientes.

Quien utilice IA deberá asumir que **el primer resultado es un borrador y no un producto terminado**.

4.4. Revisión editorial independiente

Cuando la naturaleza del producto lo justifique, deberá existir una revisión por una persona distinta de quien realizó la producción principal.

Esta segunda revisión cumple una función diferente.

Quien elaboró un texto conoce la investigación y puede, precisamente por esa familiaridad, completar mentalmente información que no se encuentra suficientemente explicada en el producto. También puede pasar por alto un error que ha visto repetidamente durante el proceso.

La revisión independiente deberá aproximarse al contenido desde la perspectiva de una persona que debe decidir si Fundamedios puede respaldar aquello que pretende publicar.

No será necesario aplicar una segunda revisión formal a toda actividad menor o rutinaria. Deberá exigirse especialmente cuando exista un nivel de riesgo que justifique un control adicional.

4.5. Factores para determinar el nivel de revisión

El grado de supervisión deberá definirse atendiendo al riesgo concreto y no únicamente al formato del producto.

Una publicación breve puede requerir más revisión que un informe extenso si contiene una acusación grave.

Deberán considerarse, entre otros, los siguientes factores:

Impacto sobre terceros. ¿Puede afectar significativamente reputación, privacidad, seguridad o derechos?

Naturaleza de las afirmaciones. ¿Contiene denuncias, acusaciones, conclusiones sobre posibles ilícitos o información especialmente sensible?

Calidad de las fuentes. ¿Existen fuentes directas y suficientemente verificadas o el producto depende de información indirecta?

Complejidad. ¿Exige interpretación jurídica, técnica, científica, estadística o especializada?

Intervención de IA. ¿La herramienta únicamente corrigió el texto o participó en análisis, selección, interpretación y redacción?

Grado de automatización. ¿Cuántas etapas del proceso dependieron de sistemas automatizados?

Confidencialidad. ¿Intervienen fuentes protegidas, datos personales sensibles o información reservada?

Reversibilidad. ¿Puede corregirse fácilmente un error o sus efectos serían difíciles de reparar una vez publicado?

Cuanto mayor sea la combinación de estos factores, mayor deberá ser el nivel de revisión.

4.6. Niveles orientativos de control

Para facilitar la aplicación práctica del Manual, Fundamedios podrá trabajar con tres niveles generales de revisión.

Nivel ordinario

Aplicable a contenidos de bajo riesgo y usos limitados de IA, como correcciones de estilo, adaptaciones sencillas o contenidos construidos sobre información pública claramente identificable.

Requerirá revisión de la persona responsable antes de publicación conforme al flujo editorial ordinario.

Nivel reforzado

Aplicable cuando la IA haya intervenido sustancialmente en la redacción o análisis, cuando existan múltiples fuentes, interpretaciones relevantes, afirmaciones susceptibles de controversia o información cuya presentación incorrecta pueda producir consecuencias significativas.

Requerirá una revisión sustantiva adicional conforme a la organización editorial correspondiente.

Nivel especial

Aplicable a investigaciones o productos de alto riesgo: acusaciones graves, fuentes confidenciales, información personal especialmente protegida, asuntos jurídicamente complejos, posibles riesgos para personas o publicaciones con capacidad de generar consecuencias institucionales significativas.

Además de los controles anteriores, deberá evaluarse la necesidad de revisión superior, especializada, jurídica o de seguridad.

La clasificación no deberá utilizarse mecánicamente. Si durante la revisión aparece un nuevo riesgo, el producto podrá pasar a un nivel superior.

4.7. Escalamiento de dudas

La persona encargada de revisar un producto no estará obligada a resolver por sí misma toda duda que encuentre.

Cuando aparezca una cuestión que exceda razonablemente sus conocimientos o atribuciones, deberá escalarse.

Por ejemplo:

- una duda sobre el alcance de una norma podrá requerir revisión jurídica;
- una metodología estadística compleja podrá requerir revisión técnica;
- un posible riesgo para una fuente podrá requerir consulta con quien dirige la investigación;
- una duda sobre datos personales podrá requerir el criterio de quien tenga responsabilidad sobre su protección;
- un posible incidente tecnológico podrá requerir revisión de seguridad.

La supervisión humana no significa que una sola persona deba saberlo todo. Significa que **las decisiones relevantes deben llegar a las personas adecuadas antes de publicar**.

4.8. Revisión jurídica

La utilización de inteligencia artificial no hará obligatoria una revisión jurídica de todos los productos.

Sin embargo, deberá considerarse cuando el contenido presente riesgos jurídicos significativos.

Podrán justificarla, entre otras circunstancias:

a) imputaciones de posibles delitos o infracciones; b) acusaciones capaces de afectar gravemente el honor o reputación; c) tratamiento de información privada o especialmente protegida; d) publicación de documentos cuya difusión pueda encontrarse restringida; e) identificación de víctimas, menores o fuentes protegidas; f) interpretaciones jurídicas determinantes para la conclusión del producto; g) existencia de medidas judiciales o administrativas relevantes; h) **amenazas o reclamaciones jurídicas vinculadas con la publicación;** o **i) cualquier situación en la que exista una duda razonable sobre la licitud de**

publicar determinada información.

La revisión jurídica deberá conocer, cuando resulte necesario, qué papel tuvo la IA en el proceso y cuáles fueron las fuentes originales utilizadas.

4.9. Verificación de afirmaciones centrales

La revisión final deberá prestar atención especial a las afirmaciones sobre las cuales descansa el producto.

No será suficiente verificar una muestra de detalles secundarios si la conclusión principal no ha sido directamente comprobada.

En una investigación, deberán identificarse las afirmaciones que, si fueran incorrectas, modificarían sustancialmente el sentido de la publicación.

Esas afirmaciones deberán revisarse contra sus fuentes.

Podrá utilizarse una lógica sencilla:

Si esta afirmación fuera falsa, ¿seguiría siendo esencialmente correcta la publicación?

Si la respuesta es no, se trata de una afirmación central y merece una verificación reforzada.

4.10. Revisión de citas

Toda cita textual relevante deberá verificarse contra su fuente original antes de la publicación.

La persona revisora deberá comprobar:

literalidad — atribución — contexto — identidad del interlocutor — y, cuando corresponda, traducción.

Cuando una cita provenga de una transcripción automatizada, deberá comprobarse en el audio o video original si se encuentra disponible.

Cuando provenga de un documento, deberá comprobarse directamente en ese documento.

No deberá utilizarse como única fuente de verificación una respuesta posterior de la misma IA que generó o extrajo inicialmente la cita.

4.11. Revisión de enlaces y referencias

Los enlaces, referencias bibliográficas, números de expedientes, identificadores de documentos, normas, sentencias y demás elementos de referencia generados o incorporados mediante IA deberán comprobarse cuando formen parte del producto.

Una referencia aparentemente precisa puede ser completamente inexistente.

Cuando un enlace sea utilizado como fuente, deberá comprobarse que conduce efectivamente al contenido que pretende respaldar y que ese contenido sostiene la afirmación correspondiente.

4.12. Revisión de cifras

Cuando las cifras constituyan un elemento relevante del producto, deberán verificarse separadamente del texto narrativo.

Esto resulta especialmente importante cuando la IA haya:

- extraído datos de documentos;
- realizado cálculos;
- convertido monedas o unidades;
- calculado porcentajes;
- comparado períodos;
- agrupado categorías; o
- producido estadísticas.

En productos de riesgo elevado, deberá conservarse la posibilidad de reconstruir el cálculo o identificar la base de datos de la cual provino.

4.13. Revisión de nombres, cargos y fechas

Antes de publicación deberán comprobarse los nombres propios, cargos, instituciones y fechas cuya precisión sea relevante.

La revisión deberá considerar especialmente si la información corresponde al **momento de los hechos**.

Una persona puede ocupar actualmente un cargo diferente del que tenía durante el acontecimiento investigado.

La IA tiende a utilizar información contextual sin distinguir siempre correctamente estos momentos temporales. La revisión humana deberá hacerlo.

4.14. Revisión del derecho a la versión

Cuando corresponda procurar la versión de una persona o institución afectada, la revisión deberá comprobar que el procedimiento se haya cumplido antes de autorizar la publicación.

Deberá revisarse:

qué se preguntó;

si las preguntas reflejaban adecuadamente los señalamientos relevantes;

qué respuesta se recibió;

si el texto la representa correctamente; y

si existen elementos sustanciales de esa respuesta que deban incorporarse para evitar una presentación incompleta.

Cuando no se haya recibido respuesta, deberá comprobarse que la forma de expresarlo sea exacta.

4.15. Revisión del tratamiento de información personal y confidencial

Antes de publicar deberá realizarse una última comprobación sobre la información cuya exposición pueda afectar a terceros.

La revisión deberá preguntar si todos los datos personales incluidos son realmente necesarios para la finalidad del producto.

También deberá comprobarse que no se hayan incorporado accidentalmente:

datos de contacto; direcciones; números de identificación; metadatos; nombres de fuentes protegidas; información contenida en capturas; referencias internas; comentarios de edición; instrucciones dirigidas a la IA; o cualquier otro elemento que no estuviera destinado al público.

Esta revisión deberá extenderse a archivos adjuntos, imágenes, documentos descargables y material audiovisual.

4.16. Revisión del titular y elementos de distribución

El titular deberá aprobarse después de conocer la versión final del producto.

La misma revisión deberá aplicarse, cuando corresponda, a subtítulos, sumarios, publicaciones para redes sociales, notificaciones, miniaturas, pies de fotografía y demás elementos utilizados para distribuir el contenido.

No deberá permitirse que una automatización genere y publique estos elementos sin control cuando puedan modificar sustancialmente la forma en que el público interpreta la información.

Una publicación puede ser correcta y su promoción ser engañosa. Ambos componentes forman parte del proceso editorial.

4.17. Uso de IA como segundo revisor

Antes de la aprobación humana podrá utilizarse una herramienta de IA para realizar una revisión adicional del producto.

Cuando resulte útil, podrá recibir instrucciones para asumir una función deliberadamente crítica:

Identificar afirmaciones que no estén respaldadas por los documentos proporcionados; señalar citas cuya fuente no pueda localizarse; detectar contradicciones entre fechas o cifras; identificar lenguaje que convierta acusaciones en hechos; señalar conclusiones que excedan la evidencia; y advertir sobre información personal o confidencial cuya inclusión deba revisarse.

Este control puede aumentar la capacidad de detección de errores, pero no modifica la responsabilidad del revisor humano.

La respuesta “no encontré problemas” producida por una herramienta **no constituirá autorización para publicar**.

4.18. Independencia de la revisión en productos de alto riesgo

En productos de alto riesgo deberá procurarse que al menos una de las revisiones sustantivas sea realizada por una persona que no haya participado directamente en todas las etapas de producción.

La independencia permite cuestionar supuestos que pueden haberse normalizado durante una investigación prolongada.

Cuando sea posible, el revisor deberá recibir acceso a las fuentes necesarias para comprobar las afirmaciones centrales y no únicamente al texto terminado.

Su función no será reescribir el producto según sus preferencias, sino determinar si aquello que Fundamedios pretende publicar se encuentra suficientemente sustentado y adecuadamente presentado.

4.19. Correcciones solicitadas durante la revisión

Cuando un revisor identifique un problema, la corrección podrá realizarse manualmente o mediante nueva asistencia de IA.

Si se utiliza nuevamente inteligencia artificial, la modificación deberá ser revisada.

Una corrección solicitada a la herramienta puede introducir un error diferente.

Por ejemplo, pedir que “corrija la atribución del tercer párrafo” puede provocar cambios adicionales en el texto. Cuando el sistema regenere una sección completa, deberá revisarse la sección completa y no únicamente la frase que motivó la modificación.

En productos sensibles deberá evitarse realizar regeneraciones amplias inmediatamente antes de publicación sin una nueva lectura integral.

4.20. Lista de comprobación previa a publicación

Para contenidos que hayan utilizado IA de manera sustantiva, la revisión final podrá apoyarse en una lista de comprobación institucional.

Sin necesidad de convertirla en un trámite formal para toda publicación, deberá permitir verificar al menos:

FUENTES ¿Las afirmaciones centrales tienen respaldo identificable?

HECHOS ¿Nombres, fechas, cifras y cargos relevantes fueron comprobados?

CITAS ¿Las citas fueron contrastadas con el original?

ATRIBUCIÓN ¿Las denuncias, opiniones y versiones continúan correctamente atribuidas?

CONTEXTO ¿La reducción o reescritura eliminó información necesaria?

CONTRASTE ¿Se incorporó o procuró la versión correspondiente cuando era necesaria?

DATOS Y CONFIDENCIALIDAD ¿Existe información que no debería hacerse pública?

IA ¿El texto contiene información que apareció durante la generación y que no estaba en los insumos validados?

TITULAR ¿Se encuentra completamente respaldado por el producto?

APROBACIÓN ¿La persona competente ha revisado y autorizado la versión que efectivamente será publicada?

Para productos de mayor riesgo podrá quedar registro de esta comprobación.

4.21. Versión aprobada y bloqueo de modificaciones posteriores

Una vez aprobada la versión definitiva, deberá evitarse que una herramienta de IA realice modificaciones sustantivas automáticas antes de la publicación sin una nueva revisión.

Esto incluye modificaciones aparentemente menores que puedan afectar significado, tales como resumir automáticamente para ajustar extensión, generar un nuevo titular, sustituir expresiones o adaptar el texto a otro formato.

La versión que se publica deberá ser la versión que fue aprobada.

Si existe una modificación sustantiva posterior a la aprobación, deberá volver al nivel de revisión que corresponda.

4.22. Publicación automatizada

Fundamedios podrá utilizar herramientas tecnológicas para programar o ejecutar técnicamente la publicación de contenidos que ya hayan sido aprobados.

Debe distinguirse entre **automatizar el acto técnico de publicación** y **automatizar la decisión editorial de publicar**.

Lo primero podrá ser permitido.

Lo segundo no deberá realizarse respecto de contenidos sustantivos sujetos a los controles de este Manual.

Por tanto, una herramienta podrá recibir una versión previamente aprobada y publicarla a una hora determinada. No deberá recibir información, producir autónomamente una nota y decidir por sí misma que se encuentra lista para hacerla pública.

4.23. Identificación de la persona que aprueba

Los productos deberán seguir los mecanismos ordinarios de aprobación editorial de Fundamedios.

Cuando el nivel de riesgo o el grado de intervención de IA lo justifique, deberá poder identificarse quién autorizó la versión final.

Esta identificación podrá realizarse mediante los propios sistemas editoriales, registros de versiones, flujos de aprobación u otros mecanismos razonables.

No será necesario crear documentación adicional cuando los sistemas existentes ya permitan reconstruir adecuadamente esta información.

4.24. Aprobación no significa transferencia absoluta de responsabilidad

La existencia de una aprobación final no elimina las responsabilidades correspondientes a las etapas anteriores.

Quien investigó deberá haber aplicado los controles de investigación.

Quien utilizó IA para redactar deberá haber revisado el resultado.

Quien verificó deberá haber realizado las comprobaciones que correspondían.

Quien aprobó deberá haber ejercido efectivamente su función de revisión.

El sistema deberá permitir determinar qué responsabilidad correspondía a cada etapa sin asumir que la última persona que presionó “publicar” responde por todo el proceso o que las actuaciones anteriores dejan de ser relevantes.

4.25. Duda razonable antes de publicar

Cuando durante la revisión final exista una duda material que no pueda resolverse razonablemente antes de la publicación, deberá evaluarse si el contenido puede publicarse sin ese elemento, si corresponde expresar la incertidumbre o si resulta necesario posponer la publicación hasta completar la verificación.

La presión de tiempo no convierte información no verificada en información publicable.

En contenidos de actualidad puede ser preferible informar con precisión aquello que se conoce y explicar qué permanece pendiente antes que utilizar IA para completar rápidamente lo que todavía no se ha podido establecer.

4.26. Registro de la intervención de IA

Fundamedios deberá poder identificar, cuando resulte relevante para control interno, si la inteligencia artificial tuvo una intervención sustantiva en la producción de determinado contenido.

No será necesario registrar como utilización sustantiva cada corrección ortográfica, búsqueda auxiliar o ajuste menor.

Sí deberá considerarse relevante cuando la IA haya intervenido significativamente en:

análisis de información; elaboración del borrador; síntesis de múltiples fuentes; extracción de datos centrales; formulación de conclusiones preliminares; producción de material visual o audiovisual; o cualquier otra actividad que haya influido materialmente en el contenido final.

Este registro permitirá aplicar controles proporcionales, investigar incidentes y evaluar posteriormente el funcionamiento de las herramientas utilizadas.

4.27. Supervisión como proceso y no como firma final

La supervisión humana deberá existir a lo largo del proceso y no únicamente al terminarlo.

Un sistema adecuadamente supervisado permite que las personas intervengan cuando:

se define la investigación; se seleccionan las fuentes; se validan los datos; se preparan los insumos; se revisa la generación; se contrastan las afirmaciones; se adoptan decisiones editoriales; y se autoriza la publicación.

No todas estas funciones deberán ser realizadas por personas diferentes. En trabajos sencillos una misma persona podrá cumplir varias de ellas.

Lo esencial será que las decisiones críticas permanezcan bajo control humano y puedan ser reconstruidas cuando resulte necesario.

4.28. Regla final de supervisión y aprobación

La inteligencia artificial puede realizar una parte importante del trabajo operativo necesario para producir un contenido, pero la capacidad técnica para realizar una tarea no equivale a autoridad para adoptar la decisión editorial correspondiente.

La regla general será:

Toda publicación sustantiva asistida por inteligencia artificial deberá llegar al público porque una persona autorizada determinó que podía publicarse, y no simplemente porque una herramienta determinó que estaba terminada.

---

5. Transparencia y divulgación sobre el uso de inteligencia artificial

La transparencia constituye uno de los principios centrales para el uso responsable de inteligencia artificial dentro de Fundamedios. Sin embargo, transparencia no significa que toda utilización de una herramienta de IA deba ser anunciada de manera individual en cada contenido publicado.

La inteligencia artificial puede intervenir de formas muy diferentes dentro de un proceso de trabajo. Puede utilizarse para corregir ortografía, transcribir una entrevista, organizar documentos, comparar información, elaborar un primer borrador, analizar una base de datos o generar una imagen. Equiparar todos estos usos y exigir idéntica divulgación produciría información poco útil para el público y podría terminar convirtiendo la transparencia en una formalidad.

El criterio deberá ser, por tanto, **la relevancia material que haya tenido la inteligencia artificial en la construcción del contenido que recibe el público y la posibilidad de que su intervención modifique razonablemente la forma en que ese contenido debe ser interpretado**.

La transparencia tendrá además una dimensión interna. Fundamedios deberá poder determinar qué herramientas utiliza, para qué finalidades, bajo qué controles y con qué intervención humana, incluso en aquellos casos en que no resulte necesario informar individualmente al público.

5.1. Transparencia institucional

Fundamedios deberá mantener una política pública general sobre el uso de inteligencia artificial en sus actividades.

Esta política deberá explicar, en lenguaje comprensible, que la organización puede utilizar herramientas de IA como apoyo para determinadas tareas de investigación, procesamiento, análisis, producción y edición de contenidos, bajo procedimientos de verificación y supervisión humana.

La declaración institucional deberá evitar dos extremos.

No deberá sugerir que Fundamedios no utiliza inteligencia artificial cuando efectivamente forma parte de sus procesos. Tampoco deberá transmitir la impresión de que los productos son generados autónomamente por sistemas automatizados cuando existe investigación, criterio profesional, revisión y aprobación humana.

La transparencia deberá describir razonablemente **cómo se integra la tecnología al trabajo institucional**.

5.2. Transparencia interna y transparencia pública

Deberán distinguirse dos niveles de información.

Transparencia interna implica que Fundamedios pueda reconstruir el uso relevante de inteligencia artificial dentro de sus propios procesos: herramienta utilizada, finalidad, intervención realizada, información proporcionada, controles aplicados y aprobación humana cuando corresponda.

Transparencia pública implica informar a lectores, usuarios o terceros cuando la intervención de IA tenga una importancia suficiente para que conocerla sea relevante para interpretar el producto.

No toda utilización registrada internamente deberá aparecer públicamente.

Esta distinción permitirá mantener una trazabilidad amplia sin llenar las publicaciones de declaraciones que no aporten información significativa al lector.

5.3. Usos que normalmente no requieren divulgación individual

Como regla general, no será necesario incorporar una advertencia específica en una publicación cuando la IA haya cumplido exclusivamente funciones instrumentales que no alteren sustancialmente su contenido.

Podrán encontrarse en esta categoría:

- corrección ortográfica y gramatical;
- mejoras menores de sintaxis;
- ajustes de formato;
- organización interna de documentos;
- generación de índices;
- clasificación auxiliar de archivos;
- transcripción posteriormente comprobada;
- traducción auxiliar revisada por una persona;

- búsqueda preliminar cuyos resultados fueron posteriormente verificados;
- propuestas de estructura que no incorporen información nueva; y
- otras tareas equivalentes sometidas a control humano.

En estos casos, la declaración institucional general sobre uso de IA podrá proporcionar transparencia suficiente.

La decisión podrá variar cuando una característica específica del contenido haga materialmente relevante explicar la intervención tecnológica.

5.4. Intervención sustantiva

Deberá evaluarse la necesidad de una divulgación específica cuando la IA haya tenido una intervención sustantiva en aquello que finalmente recibe el público.

Esto podrá ocurrir cuando la herramienta haya participado de manera significativa en:

- **la generación del texto principal;**
- **el análisis de grandes conjuntos de datos que sustenten las conclusiones;**
- **la identificación de patrones determinantes para una investigación;**
- **la generación o modificación sustancial de imágenes, audio o video;**
- **una metodología automatizada novedosa cuyo funcionamiento sea necesario para comprender los resultados;** o
- **cualquier otra actividad en la que ocultar la intervención de IA pueda producir una percepción materialmente equivocada sobre cómo se obtuvo o produjo el contenido.**

La existencia de una intervención sustantiva no determinará necesariamente que todas las publicaciones requieran exactamente la misma fórmula de divulgación. La forma y ubicación deberán responder al contexto.

5.5. Divulgación metodológica

En investigaciones en las que la inteligencia artificial haya sido utilizada para procesar o analizar grandes cantidades de información, la transparencia podrá incorporarse dentro de la explicación metodológica.

Por ejemplo, si una herramienta fue utilizada para clasificar miles de documentos, identificar determinadas categorías o realizar una primera selección de registros posteriormente revisada por personas, podrá resultar relevante explicar esa metodología.

La divulgación deberá permitir comprender:

qué función realizó la herramienta;

sobre qué conjunto de información trabajó;

qué controles humanos se aplicaron; y

qué limitaciones relevantes tuvo el procedimiento.

No será necesario divulgar prompts completos, configuraciones internas, información confidencial o detalles técnicos cuya publicación no contribuya razonablemente a comprender la metodología.

5.6. Contenido visual generado mediante IA

Cuando una imagen haya sido generada íntegramente mediante inteligencia artificial y se utilice en una publicación, deberá evaluarse especialmente la posibilidad de que el público la interprete como una representación documental de hechos reales.

Cuando exista esa posibilidad, la naturaleza artificial de la imagen deberá identificarse claramente.

Podrán utilizarse expresiones como:

“Imagen generada mediante inteligencia artificial.”

“Ilustración generada mediante inteligencia artificial.”

o una fórmula equivalente apropiada al contexto.

La identificación deberá ser suficientemente visible para impedir una interpretación razonablemente equivocada sobre el origen del material.

5.7. Modificación de imágenes reales

Deberá distinguirse entre ajustes técnicos y modificaciones sustantivas.

Correcciones ordinarias de tamaño, encuadre, formato, compresión, iluminación u otros ajustes que no alteren el significado documental de una imagen no deberán tratarse necesariamente como generación mediante IA.

En cambio, cuando una herramienta:

- incorpore o elimine personas u objetos;
- modifique el escenario;
- reconstruya partes relevantes inexistentes;
- altere elementos capaces de cambiar la interpretación de la fotografía; o
- produzca cualquier otra modificación sustantiva,

deberá identificarse adecuadamente la alteración cuando el material vaya a ser publicado.

En contenidos periodísticos, una imagen modificada sustancialmente no deberá presentarse de manera que pueda ser confundida con evidencia fotográfica original.

5.8. Audio y video sintéticos

La generación o modificación mediante IA de voces, declaraciones o representaciones audiovisuales de personas reales requiere un estándar especialmente estricto.

No deberá utilizarse material sintético de manera que una persona razonable pueda creer que alguien pronunció palabras, realizó acciones o participó en acontecimientos que no ocurrieron.

Cuando exista una finalidad legítima para utilizar una recreación, simulación o contenido sintético, su naturaleza deberá indicarse de manera clara y próxima al contenido.

La divulgación no deberá ocultarse exclusivamente en términos generales, metadatos o páginas separadas cuando el propio material pueda inducir inmediatamente a una interpretación equivocada.

5.9. Recreaciones e ilustraciones

Podrán utilizarse herramientas generativas para elaborar recursos conceptuales o ilustrativos que ayuden a explicar un asunto.

En estos casos deberá existir una separación clara entre **ilustración** y **evidencia**.

Una recreación visual de una situación descrita en una investigación no deberá presentarse como fotografía del acontecimiento.

Cuando exista riesgo de confusión deberá identificarse como ilustración, recreación o contenido generado, según corresponda.

Cuanto mayor sea el realismo del material y mayor su proximidad a acontecimientos o personas reales, mayor deberá ser la claridad de la identificación.

5.10. Divulgación de texto asistido por IA

No deberá establecerse como regla que todo texto en cuya elaboración haya participado una herramienta de inteligencia artificial lleve automáticamente una etiqueta de “generado por IA”.

Una declaración de ese tipo puede resultar imprecisa cuando el contenido fue investigado, verificado, modificado y aprobado por personas.

Cuando corresponda realizar una divulgación específica, deberá describirse la función real de la herramienta.

Por ejemplo:

Para el procesamiento inicial de la documentación utilizada en esta investigación se emplearon herramientas de inteligencia artificial. Los resultados fueron posteriormente revisados y verificados por el equipo de Fundamedios antes de su utilización.

O, cuando la intervención haya sido diferente:

Este contenido fue elaborado con asistencia de herramientas de inteligencia artificial para la organización y redacción inicial de la información. Su contenido factual fue revisado y aprobado por el equipo responsable antes de la publicación.

La fórmula deberá adaptarse al uso efectivamente realizado.

5.11. Prohibición de disclosures engañosos

Una declaración de transparencia también puede ser engañosa.

No deberá utilizarse una fórmula genérica que minimice una intervención sustancial.

Si una IA realizó un análisis determinante para las conclusiones de una investigación, no sería suficiente describir su participación únicamente como “apoyo editorial”.

De la misma manera, no deberá afirmarse que un contenido fue “generado por IA” cuando la herramienta únicamente corrigió ortografía.

La divulgación deberá guardar correspondencia con el proceso real.

5.12. La divulgación no sustituye la verificación

Informar que se utilizó inteligencia artificial no reduce el estándar de exactitud de Fundamedios.

Una advertencia como “este contenido fue generado con IA” no convierte en aceptables las alucinaciones, citas incorrectas, datos falsos o falta de supervisión.

La transparencia explica **cómo se produjo** un contenido. No exonera de responsabilidad sobre **qué se publicó**.

En consecuencia:

el disclosure es un mecanismo de transparencia, no una cláusula de exoneración.

5.13. Transparencia sobre limitaciones metodológicas

Cuando la utilización de IA introduzca limitaciones relevantes para interpretar los resultados, estas deberán explicarse cuando corresponda.

Por ejemplo, si un análisis automatizado se realizó sobre una muestra incompleta, si una clasificación requirió criterios probabilísticos o si existieron documentos que el sistema no pudo procesar adecuadamente, estas circunstancias podrán ser necesarias para comprender el alcance de las conclusiones.

No será necesario describir cada limitación técnica imaginable.

Deberán comunicarse aquellas que puedan modificar razonablemente la interpretación del resultado.

5.14. Información sobre las herramientas utilizadas

No será necesario identificar públicamente por nombre comercial cada herramienta utilizada en actividades ordinarias.

Podrá ser conveniente hacerlo cuando:

la herramienta constituya parte relevante de la metodología;

su funcionamiento tenga relación directa con las conclusiones;

la publicación evalúe precisamente el desempeño de esa tecnología;

la identificación resulte necesaria para reproducir o comprender el procedimiento; o

exista otra razón editorial suficiente.

Internamente sí deberá existir conocimiento adecuado de las herramientas institucionalmente autorizadas.

5.15. Transparencia frente a fuentes y colaboradores

Cuando una persona proporcione información bajo condiciones específicas de confidencialidad y se pretenda introducir esa información en un sistema de inteligencia artificial, deberán respetarse las condiciones establecidas en los capítulos anteriores.

Cuando la naturaleza de la relación, el compromiso asumido o el riesgo para la fuente haga materialmente relevante conocer que su información será procesada mediante determinados sistemas, deberá evaluarse la necesidad de informarlo previamente.

La transparencia con una fuente no deberá convertirse en una autorización genérica para introducir información protegida en cualquier herramienta.

Los límites institucionales continuarán aplicándose aunque la fuente manifieste no tener objeciones.

5.16. Solicitudes del público sobre el uso de IA

Cuando exista una consulta razonable sobre la utilización de inteligencia artificial en una publicación específica, Fundamedios deberá procurar responder de forma consistente con este Manual y con la información que pueda legítimamente divulgar.

La respuesta podrá explicar el tipo de intervención realizada y los controles humanos aplicados.

No existirá obligación de revelar información que pueda comprometer:

fuentes; investigaciones en curso; información confidencial; seguridad; datos personales; sistemas internos; credenciales; prompts reservados cuya divulgación genere riesgos; o derechos de terceros.

La transparencia deberá coexistir con las obligaciones de protección de información.

5.17. Correcciones relacionadas con el uso de IA

Cuando un error publicado tenga relación material con el uso de inteligencia artificial, la política general de correcciones de Fundamedios continuará siendo aplicable.

Si para comprender la naturaleza del error resulta relevante conocer que una herramienta produjo, interpretó o procesó incorrectamente determinada información, esta circunstancia podrá explicarse.

Sin embargo, no deberá presentarse la IA como sujeto responsable del error.

Será preferible describir concretamente lo ocurrido.

Por ejemplo, en lugar de:

“La inteligencia artificial se equivocó”,

podrá explicarse, cuando corresponda:

“Durante el procesamiento automatizado de la información se produjo una identificación incorrecta que no fue detectada durante la revisión previa a publicación.”

Esta formulación permite comprender tanto el origen tecnológico como la falla del control humano.

5.18. Transparencia sobre automatización

Cuando Fundamedios implemente sistemas capaces de producir automáticamente contenidos recurrentes a partir de datos estructurados, deberá evaluarse si el público necesita conocer esa automatización.

La decisión dependerá de la naturaleza del contenido y de la intervención humana existente.

Cuando un sistema genere actualizaciones automáticas con mínima intervención humana, la divulgación deberá ser más clara que cuando la automatización únicamente prepare un borrador posteriormente revisado y aprobado.

En cualquier caso, las reglas de supervisión del apartado anterior continuarán aplicándose a los contenidos sustantivos.

5.19. Declaración institucional sobre uso responsable de inteligencia artificial

Fundamedios deberá contar con una declaración pública permanente, accesible desde sus plataformas institucionales, que explique de forma general su política sobre inteligencia artificial.

Una fórmula institucional podrá desarrollarse sobre la siguiente base:

Fundamedios utiliza herramientas de inteligencia artificial como apoyo para determinadas actividades de investigación, procesamiento de información, análisis y producción de contenidos. Su utilización se encuentra sujeta a procedimientos internos de seguridad, protección de datos, verificación de fuentes y supervisión humana.

La inteligencia artificial no sustituye la responsabilidad editorial de Fundamedios. Las decisiones sobre la verificación, interpretación, aprobación y publicación de contenidos corresponden a personas responsables dentro de la organización.

Cuando la intervención de inteligencia artificial sea material para comprender cómo fue producido un contenido —particularmente en la generación o modificación sustancial de material visual o audiovisual o en metodologías automatizadas relevantes— Fundamedios procurará identificar dicha intervención de manera clara y proporcional.

Esta declaración podrá adaptarse posteriormente a la política editorial y al lenguaje institucional definitivo.

5.20. Declaración interna de uso responsable

Separadamente de la declaración dirigida al público, toda persona que tenga acceso institucional a herramientas de inteligencia artificial deberá conocer y aceptar las reglas aplicables a su utilización.

Esta declaración interna deberá vincularse con las obligaciones desarrolladas a lo largo del Manual.

Podrá comprender, entre otros compromisos:

I. Uso institucional. Utilizar las cuentas, perfiles, herramientas y entornos autorizados por Fundamedios para las actividades institucionales.

II. Protección de información. No introducir datos personales, información confidencial, fuentes protegidas o documentación restringida fuera de los entornos autorizados.

III. Verificación. No utilizar como hecho publicable información cuya única base sea una generación de inteligencia artificial.

IV. Supervisión. Revisar los resultados producidos mediante IA antes de incorporarlos a productos institucionales.

V. Trazabilidad. Conservar los registros exigidos cuando la intervención de la herramienta resulte sustantiva.

VI. Seguridad. Proteger las credenciales y respetar los controles de acceso, almacenamiento e integración.

VII. Incidentes. Comunicar oportunamente errores, exposiciones, usos indebidos o incidentes relacionados con estas herramientas.

VIII. Responsabilidad. No utilizar inteligencia artificial para eludir los procedimientos editoriales, jurídicos, de protección de datos o de seguridad establecidos por Fundamedios.

Esta declaración deberá ser aceptada como condición para el acceso a los entornos institucionales de IA cuando Fundamedios implemente el sistema.

5.21. Información que no deberá divulgarse por razones de transparencia

La obligación de transparencia no implica publicar la arquitectura completa de seguridad, instrucciones internas, credenciales, configuraciones sensibles, bases de conocimiento, información sobre vulnerabilidades o procedimientos cuya divulgación pueda facilitar ataques o extracción de información.

Tampoco deberán publicarse prompts que contengan información protegida simplemente para demostrar cómo se utilizó la herramienta.

Fundamedios deberá poder documentar internamente sus procesos sin convertir esa documentación en información pública cuando existan razones legítimas para protegerla.

5.22. Una declaración no corrige una práctica incorrecta

La transparencia deberá reflejar un procedimiento que realmente exista.

No será suficiente publicar una política de inteligencia artificial si, en la práctica, las personas utilizan cuentas personales, introducen información sin clasificación, omiten verificaciones o publican resultados sin supervisión.

La declaración pública será la expresión externa de un sistema interno de gestión.

Por ello, antes de comunicar estándares de utilización de IA al público, Fundamedios deberá procurar que existan las condiciones institucionales necesarias para cumplirlos.

5.23. Aplicación a las distintas iniciativas de Fundamedios

Los principios de transparencia deberán aplicarse de manera coherente a los productos elaborados bajo las diferentes iniciativas, proyectos, plataformas y denominaciones vinculadas con Fundamedios, incluyendo Ecuador Chequea y Liberties 360.

La forma específica de divulgación podrá variar según la naturaleza del producto.

Una metodología de fact-checking puede requerir una explicación diferente de la utilizada para una ilustración generada, una investigación documental o un contenido institucional.

La diversidad de formatos no deberá producir estándares contradictorios.

El principio común será que una persona que reciba un contenido **no sea inducida a una percepción materialmente equivocada sobre el origen, autenticidad o proceso de elaboración de aquello que está observando**.

5.24. Criterio para decidir si corresponde una divulgación específica

Cuando exista duda sobre si una utilización de inteligencia artificial debe identificarse públicamente, podrá aplicarse la siguiente pregunta:

¿Conocer que la inteligencia artificial realizó esta función cambiaría razonablemente la manera en que el público interpreta, evalúa o confía en este contenido?

Si la respuesta es claramente afirmativa, deberá favorecerse la divulgación.

También deberá favorecerse cuando exista riesgo de que material sintético sea confundido con evidencia real.

Si la intervención fue puramente instrumental, estuvo plenamente sometida a revisión humana y conocerla no aporta información material para comprender el producto, la declaración institucional general podrá resultar suficiente.

5.25. Regla final de transparencia

La transparencia deberá ser **útil, proporcional y verdadera**.

No deberá convertirse en una sucesión automática de etiquetas que termine ocultando los usos verdaderamente relevantes de inteligencia artificial. Tampoco deberá utilizarse una declaración institucional general para evitar explicar una intervención tecnológica que sea material para comprender una publicación determinada.

La regla general será:

Fundamedios deberá informar sobre el uso de inteligencia artificial cuando esa información sea relevante para comprender cómo se obtuvo, produjo o representó un contenido, y deberá hacerlo describiendo la intervención real de la herramienta y la supervisión humana aplicada.

---

6. Publicación, archivo del producto y trazabilidad posterior

La publicación no constituye el final absoluto del proceso de utilización de inteligencia artificial. Una vez que un contenido llega al público, Fundamedios deberá conservar, en la medida razonablemente necesaria, los elementos que permitan conocer cómo se construyó, qué información lo sustentó, qué intervención tuvo la inteligencia artificial y qué controles humanos fueron aplicados.

Esta fase cumple varias funciones simultáneas. Permite responder ante cuestionamientos sobre una publicación, realizar correcciones, actualizar contenidos, defender la metodología utilizada, investigar posibles incidentes y aprovechar la experiencia adquirida para mejorar futuros procesos.

La trazabilidad posterior no significa conservar indefinidamente todas las conversaciones, borradores o respuestas generadas durante una investigación. El objetivo será mantener **evidencia suficiente del proceso**, aplicando al mismo tiempo los principios de minimización, protección de datos y conservación limitada desarrollados en este Manual.

6.1. Correspondencia entre la versión aprobada y la versión publicada

La versión que llegue al público deberá corresponder con aquella que haya completado el procedimiento de revisión y aprobación.

Una vez aprobada, no deberán introducirse modificaciones sustantivas mediante inteligencia artificial sin someterlas nuevamente al control correspondiente.

Esto adquiere particular importancia cuando existen sistemas que adaptan automáticamente un contenido para distintos canales. Una nota aprobada para el portal puede posteriormente transformarse en un resumen, una publicación para redes sociales, un boletín o una descripción audiovisual. Cada transformación deberá conservar el sentido del producto original.

La aprobación de un contenido **no equivale automáticamente a la aprobación de todas las versiones que una IA pueda generar a partir de él**.

6.2. Control inmediatamente anterior a la publicación

Cuando la naturaleza del producto lo justifique, deberá realizarse una última comprobación sobre la versión que efectivamente será publicada.

Esta revisión no pretende repetir todo el procedimiento editorial. Su finalidad será detectar modificaciones accidentales producidas durante la carga, edición, adaptación o programación del contenido.

Deberá prestarse especial atención a:

- **titular y subtítulos;**
- **nombres y cargos;**
- **cifras y fechas centrales;**
- **enlaces;**
- **imágenes y pies de fotografía;**
- **atribuciones;**
- **elementos gráficos;**
- **documentos adjuntos;**
- **etiquetas de contenido generado o modificado mediante IA cuando correspondan;** y
- **cualquier información que haya sido incorporada después de la aprobación principal.**

En publicaciones de alto riesgo, esta comprobación deberá alcanzar también la correspondencia entre las afirmaciones centrales y la evidencia que las sustenta.

6.3. Registro mínimo del producto

Cuando la intervención de inteligencia artificial haya sido sustantiva, deberá conservarse un registro suficiente para reconstruir razonablemente su utilización.

El registro podrá formar parte del propio expediente digital de la investigación y no requerirá necesariamente un sistema independiente.

Según la naturaleza del producto, podrá incluir:

1. identificación del producto publicado;
2. fecha de publicación;
3. persona o equipo responsable;
4. herramienta o entorno institucional de IA utilizado;
5. función principal para la cual se utilizó;
6. fuentes o conjunto de información utilizado como base;
7. prompts o metodología relevante cuando su conservación sea necesaria;
8. revisiones adicionales realizadas;
9. persona que otorgó la aprobación final, cuando corresponda; y
10. enlace, archivo o copia de la versión efectivamente publicada.

El nivel de detalle deberá aumentar en función del riesgo.

Una corrección lingüística no deberá generar el mismo registro que una investigación cuyo análisis principal haya dependido del procesamiento automatizado de miles de documentos.

6.4. Expediente de respaldo

Los productos de mayor relevancia deberán conservar un expediente de respaldo que permita demostrar razonablemente cómo se llegó a las afirmaciones publicadas.

No deberá confundirse el expediente con una acumulación indiscriminada de todos los archivos utilizados.

Su contenido deberá concentrarse en aquello que resulte necesario para reconstruir el trabajo.

Podrán conservarse:

fuentes originales relevantes; documentos centrales; bases utilizadas; verificaciones; comunicaciones necesarias para acreditar el contraste; respuestas de personas o instituciones consultadas; versiones sustantivas del producto; metodología empleada; y registros relevantes del uso de IA.

Cuando el producto incluya afirmaciones de alto impacto, deberá procurarse que el expediente permita identificar específicamente el respaldo de esas afirmaciones.

6.5. Archivo de fuentes originales

Cuando una fuente pueda cambiar, desaparecer o ser modificada posteriormente, deberá evaluarse la conveniencia y licitud de conservar evidencia de la versión efectivamente consultada.

Esto puede ser relevante para páginas web, publicaciones en redes sociales, bases de datos dinámicas, documentos que posteriormente sean sustituidos o información pública susceptible de actualización.

La forma de conservación dependerá del caso y podrá consistir en una copia del documento, archivo, captura, registro de fecha, referencia institucional u otro mecanismo adecuado.

La conservación deberá respetar las reglas sobre propiedad intelectual, protección de datos, confidencialidad y seguridad aplicables.

6.6. Registro de la intervención de inteligencia artificial

La trazabilidad deberá permitir distinguir entre **lo que provino de una fuente y lo que fue producido por una herramienta**.

Cuando resulte relevante, deberá poder reconstruirse una secuencia como la siguiente:

Fuente original → procesamiento mediante IA → resultado intermedio → verificación humana → contenido aprobado → publicación.

Esta distinción resulta especialmente importante cuando la herramienta haya realizado extracción de datos, transcripción, clasificación, resumen, análisis o generación de un borrador.

Si posteriormente aparece un error, esta secuencia permitirá determinar si se encontraba en la fuente original, fue introducido durante el procesamiento automatizado, surgió durante la edición humana o apareció en una modificación posterior.

6.7. Conservación de prompts

No todos los prompts deberán conservarse.

Deberán preservarse aquellos que tengan valor metodológico, probatorio o de trazabilidad.

Podrá ser relevante conservarlos cuando:

- definieron la metodología de análisis;
- establecieron criterios de clasificación;
- fueron utilizados para procesar grandes cantidades de información;
- produjeron resultados centrales para una investigación;
- permitan explicar un incidente;
- correspondan a una plantilla institucional; o
- resulten necesarios para reconstruir cómo se obtuvo una conclusión.

Las conversaciones exploratorias, reformulaciones ordinarias y consultas sin incidencia material podrán quedar fuera del expediente.

Cuando un prompt contenga información confidencial o datos personales, su conservación deberá someterse a los mismos controles aplicables al resto de esa información.

6.8. Conservación del resultado original de IA

En determinados casos será útil conservar el resultado producido originalmente por la herramienta antes de la intervención humana.

Esto permitirá distinguir qué produjo el sistema y qué modificaciones introdujo posteriormente el equipo.

No será necesario hacerlo para todas las actividades.

Deberá considerarse especialmente cuando el resultado haya tenido una función sustancial en una investigación, cuando exista un riesgo elevado, cuando se utilice una metodología experimental o cuando resulte necesario evaluar posteriormente el desempeño de la herramienta.

6.9. Producto final como registro separado

La versión definitivamente publicada deberá conservarse de manera que pueda distinguirse claramente de los borradores.

Cuando un contenido sea posteriormente actualizado, deberá procurarse que exista trazabilidad suficiente para determinar qué versión estuvo disponible al público en un momento determinado cuando ello tenga relevancia.

Esto será particularmente importante cuando una modificación afecte:

hechos centrales, acusaciones, cifras, conclusiones, nombres, atribuciones o cualquier elemento que haya dado lugar a una corrección.

El objetivo no será mantener públicamente todas las versiones anteriores, sino conservar internamente información suficiente para reconstruir cambios relevantes.

6.10. Actualizaciones posteriores

Un contenido publicado puede requerir actualización sin que la versión anterior haya sido incorrecta.

Puede aparecer nueva información, producirse una decisión judicial, responder una persona que no había contestado inicialmente o modificarse una circunstancia relevante.

Cuando la IA sea utilizada para actualizar una publicación, deberá proporcionársele claramente:

la versión vigente; la nueva información verificada; qué elementos pueden modificarse; y cuáles deben conservarse.

La herramienta no deberá regenerar innecesariamente la totalidad de una publicación cuando ello pueda introducir cambios en información que ya había sido verificada.

Cuando sea técnicamente razonable, deberá preferirse una modificación controlada de los elementos afectados.

6.11. Correcciones

Deberá distinguirse una **actualización** de una **corrección**.

Existe actualización cuando nueva información modifica o amplía un contenido que era correcto al momento de publicación.

Existe corrección cuando Fundamedios modifica información que era incorrecta, imprecisa o insuficientemente presentada.

La utilización de IA para detectar o realizar la corrección no altera esta diferencia.

Cuando corresponda explicar públicamente una corrección, deberá hacerse conforme a las políticas editoriales aplicables y con un nivel de transparencia proporcional a la importancia del error.

6.12. Corrección en productos derivados

Cuando una publicación sea corregida, deberá evaluarse si el error fue reproducido en otros productos.

Una misma información puede haber sido transformada mediante IA en:

una publicación web, una alerta, un boletín, una publicación para redes sociales, un gráfico, un video, una traducción o un documento posterior.

Corregir únicamente la fuente principal puede resultar insuficiente.

Deberá identificarse razonablemente dónde se reprodujo la información incorrecta y adoptar las medidas necesarias según la relevancia y posibilidad de corrección.

Esta obligación adquiere especial importancia en sistemas automatizados capaces de reutilizar contenidos anteriores.

6.13. Corrección de la memoria institucional

Cuando el error tenga origen en información almacenada dentro del entorno institucional de IA, la corrección deberá alcanzar también ese entorno.

No deberá permitirse que una afirmación públicamente corregida permanezca dentro de una base de conocimiento activa como información válida.

Deberá revisarse, según corresponda:

- la fuente incorporada;
- los resúmenes derivados;
- las bases estructuradas;
- las memorias del sistema;
- las plantillas que contengan el dato;
- los documentos internos que lo reproduzcan; y
- cualquier otro elemento capaz de volver a introducir el error.

La corrección deberá procurar **interrumpir la cadena de reproducción**, no solamente modificar su manifestación pública más visible.

6.14. Derecho de rectificación, actualización o supresión de información

Cuando Fundamedios reciba una solicitud relacionada con datos personales, rectificación, actualización, eliminación u otro derecho reconocido por el ordenamiento jurídico aplicable, deberá poder identificar razonablemente dónde se encuentra la información correspondiente.

La existencia de inteligencia artificial no deberá impedir el ejercicio de derechos sobre los datos.

Por esta razón, la arquitectura institucional deberá evitar, en la medida técnicamente razonable, sistemas en los cuales resulte imposible conocer dónde se almacenó determinada información o retirarla de los repositorios activos cuando jurídicamente corresponda.

Cuando la información se encuentre incorporada a productos periodísticos, archivos históricos, bases de conocimiento u otros entornos con finalidades diferentes, deberá evaluarse cada tratamiento conforme a su régimen jurídico y finalidad.

6.15. Reclamaciones y cuestionamientos

Cuando una persona cuestione una publicación producida con asistencia de IA, la primera respuesta interna no deberá consistir en volver a preguntar a la misma herramienta si considera correcto su resultado.

Deberá regresarse a la evidencia.

La revisión deberá identificar:

qué afirmación se cuestiona → qué fuente la respalda → cómo fue procesada → qué controles se realizaron → qué versión se publicó.

Solo después de reconstruir estos elementos podrá determinarse si corresponde mantener, explicar, actualizar o corregir el contenido.

La herramienta podrá asistir en el análisis de la reclamación, pero no deberá actuar como árbitro de la corrección de su propio trabajo.

6.16. Preservación ante una controversia jurídica

Cuando exista una reclamación formal, amenaza de acción jurídica, procedimiento administrativo o judicial, o una expectativa razonable de controversia relacionada con una publicación, deberá preservarse la información necesaria para la defensa y reconstrucción del proceso.

En estas circunstancias podrán suspenderse las eliminaciones ordinarias respecto de los materiales relevantes.

Deberán conservarse especialmente las fuentes, versiones, comunicaciones, verificaciones, registros de aprobación y elementos relevantes de la intervención de IA.

La preservación deberá realizarse sin ampliar innecesariamente el acceso a esa información.

6.17. Acceso a los expedientes posteriores a publicación

La publicación de un contenido no convierte automáticamente en pública toda la documentación utilizada para producirlo.

Los expedientes podrán contener fuentes confidenciales, información personal, análisis internos, documentos protegidos, estrategias editoriales o información cuya difusión genere riesgos.

Los permisos de acceso deberán mantenerse conforme a la clasificación de la información.

Cuando una investigación concluya, deberá revisarse si determinados accesos que fueron necesarios durante su desarrollo continúan estando justificados.

6.18. Reutilización de investigaciones anteriores

La información archivada podrá utilizarse posteriormente como antecedente para nuevos trabajos.

Sin embargo, su existencia dentro de un repositorio institucional no implica que continúe siendo actual.

Antes de reutilizar información anterior deberá comprobarse, según corresponda:

si continúa vigente; si fue posteriormente corregida; si aparecieron nuevos hechos; si cambió la situación jurídica; si la persona ocupa todavía el mismo cargo; si las cifras corresponden al período actual; y si las condiciones bajo las cuales fue obtenida permiten su nuevo uso.

La IA no deberá tratar automáticamente todo el archivo institucional como conocimiento actual.

6.19. Aprendizaje institucional

El archivo posterior a publicación deberá servir también para mejorar la utilización de inteligencia artificial.

Fundamedios podrá revisar periódicamente productos en los que estas herramientas hayan tenido una intervención importante para identificar:

- tareas en las que producen buenos resultados;
- errores recurrentes;
- prompts que requieren ajustes;
- verificaciones que detectan problemas con mayor frecuencia;
- herramientas que han dejado de resultar adecuadas;
- procesos que pueden simplificarse;
- controles que deben reforzarse; y
- necesidades de capacitación.

Este aprendizaje deberá incorporarse progresivamente a las metodologías, plantillas y procedimientos institucionales.

6.20. Indicadores internos

Sin convertir la actividad editorial en un sistema de medición excesivamente burocrático, Fundamedios podrá utilizar indicadores para conocer el funcionamiento de su sistema de IA.

Podrán resultar útiles, por ejemplo:

a) número y tipo de incidentes relevantes; b) errores detectados antes de publicación; c) errores detectados después de publicación; d) herramientas involucradas; e) tipos de tareas con mayor frecuencia de corrección; f) incidentes de protección de datos o seguridad;

g) **porcentaje de usuarios capacitados;** y **h) resultados de revisiones periódicas de accesos y permisos.**

Estos indicadores deberán utilizarse para mejorar el sistema y no para incentivar volumen de producción a costa de la calidad.

6.21. Auditoría del procedimiento

Fundamedios podrá realizar revisiones periódicas sobre la aplicación del Manual.

La revisión no deberá concentrarse exclusivamente en comprobar si existe documentación. Deberá evaluar si los controles funcionan en la práctica.

Podrá seleccionarse una muestra de productos y reconstruir:

> **cómo se obtuvo la información → qué herramienta se utilizó → dónde se almacenó → cómo fue verificada → cómo se produjo el contenido → quién lo revisó → qué se publicó → qué evidencia permanece disponible.**

Este ejercicio permitirá detectar diferencias entre el procedimiento escrito y la práctica real.

Cuando existan desviaciones recurrentes, deberá evaluarse si el problema se encuentra en el comportamiento de los usuarios, en la capacitación, en las herramientas disponibles o en que el propio procedimiento resulta innecesariamente complejo.

6.22. Evaluación de herramientas a partir de resultados reales

La autorización de una herramienta no deberá considerarse permanente.

La experiencia obtenida mediante su utilización podrá justificar mantenerla, restringir determinadas funciones, modificar su configuración o dejar de utilizarla.

Deberán considerarse especialmente:

frecuencia y naturaleza de errores; capacidad de protección de información; posibilidades de administración institucional; trazabilidad; controles de acceso; comportamiento de nuevas funcionalidades; condiciones de tratamiento de datos; y capacidad para integrarse adecuadamente al procedimiento de Fundamedios.

Una herramienta técnicamente avanzada puede no resultar apropiada para determinadas actividades si sus condiciones de seguridad, administración o trazabilidad son insuficientes.

6.23. Retiro de una herramienta

Cuando Fundamedios decida dejar de utilizar una herramienta de inteligencia artificial, deberá existir un procedimiento razonable de salida.

Deberá determinarse:

1. qué información institucional permanece almacenada;
2. qué información debe exportarse o conservarse;
3. qué proyectos o asistentes dependen de ella;
4. qué integraciones deben revocarse;
5. qué usuarios mantienen acceso;
6. qué información debe eliminarse cuando sea posible;
7. qué metodologías o prompts institucionales deben preservarse; y
8. qué herramienta o procedimiento sustituirá las funciones necesarias.

El retiro deberá procurar que la organización no pierda conocimiento institucional ni deje información innecesariamente expuesta en servicios que ya no utiliza.

6.24. Cierre del expediente

Cuando un proyecto haya concluido y haya transcurrido el período durante el cual su información necesita permanecer activa, deberá revisarse su expediente.

La revisión deberá determinar qué información:

permanece activa;

pasa a archivo;

debe anonimizarse;

debe conservarse por razones jurídicas, históricas o institucionales; y

puede ser eliminada.

El cierre deberá incluir, cuando corresponda, la retirada de la información de memorias activas de inteligencia artificial.

Un proyecto cerrado no deberá permanecer indefinidamente conectado a todos los sistemas de IA únicamente porque técnicamente sea posible hacerlo.

6.25. El ciclo completo de información y producción

Con este procedimiento, la utilización institucional de inteligencia artificial deberá concebirse como un ciclo y no como una sucesión de consultas aisladas a una herramienta.

El flujo general será:

> **PLANIFICAR → RECOPILAR → CLASIFICAR → VERIFICAR → ALMACENAR → ANALIZAR → PRODUCIR → REVISAR → APROBAR → PUBLICAR → CONSERVAR → CORREGIR → APRENDER**

Cada etapa deberá alimentar a la siguiente.

La información incorrecta detectada durante la revisión deberá regresar a verificación.

Una fuente nueva obtenida durante la redacción deberá regresar a investigación.

Un error detectado después de publicación deberá corregirse no solamente en el producto, sino en los repositorios que puedan reproducirlo.

Una lección obtenida de un incidente deberá regresar al procedimiento y modificarlo cuando resulte necesario.

De esta manera, el sistema no será lineal sino **retroalimentado**.

6.26. Regla de cierre del Capítulo IV

La utilización responsable de inteligencia artificial para producir contenidos no termina cuando una persona presiona el botón de publicación.

Fundamedios deberá mantener capacidad suficiente para explicar y reconstruir los aspectos relevantes del proceso que condujo al producto, corregir errores cuando aparezcan y evitar que información incorrecta vuelva a incorporarse a nuevas publicaciones.

La regla de cierre será:

Todo producto asistido sustancialmente por inteligencia artificial deberá poder ser defendido por sus fuentes, explicado por su proceso, atribuido a decisiones humanas identificables y corregido tanto en su manifestación pública como en los sistemas internos que lo hicieron posible.

CAPÍTULO

V

Gobernanza institucional del sistema de inteligencia artificial



La aplicación de este Manual requiere una estructura institucional que permita convertir sus reglas en prácticas efectivas. No será suficiente determinar cómo debe recopilarse información, qué datos pueden incorporarse a una herramienta o qué controles deben aplicarse antes de publicar si, paralelamente, no se establece quién administra el entorno tecnológico, quién autoriza herramientas, cómo ingresan y salen usuarios y quién puede modificar las configuraciones institucionales.

Fundamedios deberá administrar la inteligencia artificial como parte de su infraestructura institucional. Esto supone mantener control sobre las cuentas utilizadas, los accesos, las bases de conocimiento, los perfiles organizacionales, las herramientas autorizadas y las reglas bajo las cuales pueden emplearse.

La gobernanza deberá procurar un equilibrio. Un sistema excesivamente rígido puede provocar que las personas recurran informalmente a herramientas externas para resolver tareas cotidianas; uno excesivamente abierto puede dispersar información institucional entre cuentas y plataformas sobre las cuales Fundamedios carezca de control.

El objetivo será disponer de un entorno suficientemente seguro y ordenado que, al mismo tiempo, resulte útil para las actividades reales de la organización.

1. El entorno institucional de inteligencia artificial

Fundamedios deberá disponer de uno o varios entornos institucionales autorizados para la utilización de inteligencia artificial.

Siempre que las características de la plataforma lo permitan, estos entornos deberán encontrarse contratados, configurados o administrados directamente por la organización y diferenciados de las cuentas personales de quienes trabajan o colaboran con ella.

El principio aplicable será sencillo:

La información institucional deberá trabajarse dentro de entornos institucionales.

Esto permitirá que el conocimiento producido durante el trabajo permanezca bajo control de Fundamedios, que los accesos puedan administrarse y que la salida de una persona no implique la pérdida de investigaciones, metodologías, prompts, configuraciones o antecedentes relevantes.

1.1. Cuenta institucional y cuentas personales

Las actividades comprendidas por este Manual deberán realizarse mediante las cuentas institucionales autorizadas cuando involucren información de Fundamedios.

No deberán utilizarse cuentas personales de ChatGPT, Gemini, Claude, Copilot u otras herramientas equivalentes para almacenar, procesar o desarrollar información institucional cuando exista un entorno institucional habilitado para esa actividad.

Esta separación no responde únicamente a razones de confidencialidad.

Una investigación realizada dentro de una cuenta personal puede quedar vinculada a una persona específica, permanecer almacenada después de que termine su relación con Fundamedios, incorporarse a memorias personales, mezclarse con información ajena a la organización o resultar inaccesible para el resto del equipo.

De igual manera, una persona no deberá trasladar sistemáticamente información desde la cuenta institucional hacia su cuenta personal para continuar trabajando fuera del entorno autorizado.

1.2. Perfil organizacional

Fundamedios podrá desarrollar dentro de su entorno institucional un **perfil organizacional de inteligencia artificial** que permita a las herramientas autorizadas comprender mejor la naturaleza, metodologías y estándares de trabajo de la organización.

Este perfil podrá incorporar progresivamente información institucional útil y previamente seleccionada, como:

a) manuales y políticas internas; b) metodologías editoriales y de verificación; c) criterios de estilo; d) estructuras de productos recurrentes; e) glosarios y denominaciones institucionales; f) publicaciones previamente validadas cuando resulte apropiado;

g) plantillas de prompts; h) procedimientos de revisión; y **i) otras fuentes de conocimiento institucional autorizadas.**

El objetivo será evitar que cada persona tenga que reconstruir individualmente el contexto institucional en su propia cuenta y permitir que el conocimiento generado por la organización pueda aprovecharse de manera ordenada.

1.3. El perfil institucional no será una memoria indiscriminada

La existencia de un perfil organizacional no significa que toda información producida por Fundamedios deba incorporarse permanentemente a la memoria de una herramienta.

Deberá existir una selección.

Podrá incorporarse información que resulte estable, útil y apropiada para su reutilización. En cambio, deberá evaluarse cuidadosamente la incorporación de investigaciones en curso, información confidencial, datos personales, expedientes jurídicos, información sobre fuentes o materiales cuya utilización futura fuera de contexto pueda generar riesgos.

Por ello deberán diferenciarse, cuando la tecnología lo permita:

Conocimiento institucional general, reutilizable en múltiples actividades.

Conocimiento por área o proyecto, disponible únicamente para determinados equipos.

Información restringida, cuyo acceso requiera autorización específica.

Información temporal, necesaria mientras se desarrolla un proyecto pero que no debe incorporarse permanentemente a la memoria activa.

Archivo, conservado por razones institucionales o jurídicas pero no consultado automáticamente por la IA.

Esta arquitectura evitará que “memoria institucional” se convierta en sinónimo de “acceso universal”.

2. Administración del entorno

La administración general del entorno institucional deberá encontrarse asignada a una o más personas expresamente autorizadas.

No será necesario crear un nuevo cargo exclusivamente para esta función. Fundamedios podrá distribuir las responsabilidades entre las áreas existentes conforme a su estructura.

Sin embargo, deberán encontrarse claramente cubiertas al menos tres dimensiones:

Administración tecnológica, relacionada con cuentas, permisos, integraciones, configuraciones y seguridad.

Administración metodológica y editorial, relacionada con los usos autorizados, metodologías, estándares de producción y funcionamiento práctico del Manual.

Cumplimiento y protección de información, relacionada con protección de datos, confidencialidad, riesgos jurídicos y respuesta ante incidentes.

Una misma persona podrá asumir más de una función cuando la dimensión de la organización lo justifique.

2.1. Administrador institucional de IA

Fundamedios deberá designar la función responsable de la administración general de sus entornos de inteligencia artificial.

Entre sus tareas podrán encontrarse:

- administrar altas y bajas de usuarios;
- asignar permisos;
- controlar privilegios administrativos;
- habilitar o deshabilitar funcionalidades;
- administrar espacios institucionales;
- mantener el inventario de herramientas autorizadas;
- revisar integraciones;
- coordinar actualizaciones relevantes;
- apoyar la respuesta frente a incidentes; y
- verificar periódicamente que las configuraciones correspondan a las políticas institucionales.

La persona administradora no deberá, por el solo hecho de desempeñar esta función, tener acceso irrestricto al contenido de todas las investigaciones cuando dicho acceso no resulte necesario.

Los privilegios técnicos tampoco deberán eliminar la compartimentación de información.

2.2. Responsabilidad de los equipos editoriales y de investigación

Los responsables editoriales y de investigación deberán determinar cómo se integra la IA dentro de sus procesos de trabajo, respetando este Manual.

Les corresponderá especialmente identificar qué actividades requieren controles adicionales, cuándo una publicación necesita una segunda revisión, cuándo corresponde escalar una duda y qué metodologías deben aplicarse a los productos de sus respectivas áreas.

El administrador tecnológico podrá decidir si una herramienta es segura y se encuentra autorizada.

Eso no significa que deba decidir si una investigación está suficientemente verificada para publicarse.

De la misma manera, un editor podrá aprobar un contenido, pero no deberá modificar por su cuenta las configuraciones de seguridad de una plataforma para habilitar una función que no se encuentre autorizada.

La gobernanza requiere distinguir estas competencias.

3. Catálogo de herramientas autorizadas

Fundamedios deberá mantener un registro actualizado de las herramientas de inteligencia artificial autorizadas para actividades institucionales.

El registro no necesitará ser complejo. Deberá permitir conocer, al menos:

Elemento	Información mínima
Herramienta	Plataforma o servicio autorizado
Finalidad	Actividades para las que puede utilizarse
Usuarios	Personas o áreas autorizadas
Información permitida	Tipo de información que puede procesarse
Restricciones	Información o funciones que no pueden utilizarse
Integraciones	Sistemas institucionales a los que tiene acceso
Responsable	Persona o función que administra la herramienta

Última revisión Fecha de evaluación institucional

La inclusión de una herramienta en el catálogo no significará que pueda utilizarse para cualquier finalidad.

Una plataforma podrá estar autorizada para trabajar con información pública y, al mismo tiempo, no encontrarse autorizada para procesar información confidencial.

3.1. Incorporación de nuevas herramientas

Cuando una persona identifique una herramienta que pueda mejorar significativamente una actividad institucional, podrá solicitar su evaluación.

No deberá incorporarla unilateralmente a un proceso que implique información institucional protegida.

La evaluación deberá ser proporcional al riesgo y considerar, según corresponda:

qué problema resuelve; qué información necesita; dónde procesa esa información; qué condiciones de privacidad ofrece; si utiliza los datos para entrenamiento; cuánto tiempo los conserva; qué controles administrativos permite; qué integraciones solicita; cómo pueden eliminarse los datos; qué mecanismos de seguridad posee; y qué ocurre cuando Fundamedios deja de utilizarla.

No será necesario realizar el mismo nivel de evaluación para una herramienta que únicamente procese información pública que para una plataforma conectada a los repositorios internos de la organización.

3.2. Funcionalidades nuevas

Una herramienta autorizada puede incorporar posteriormente funciones que cambien sustancialmente su nivel de riesgo.

La aparición de una nueva función no implicará su autorización automática.

Especial atención requerirán funcionalidades que incorporen:

a) memoria persistente; b) navegación autónoma; c) acceso al correo electrónico; d) acceso a repositorios documentales; e) ejecución de código; f) control de aplicaciones; g) generación de contenido audiovisual realista; h) agentes capaces de realizar acciones; o **i) nuevas formas de compartir información con terceros.**

Cuando una funcionalidad modifique sustancialmente aquello que la herramienta puede conocer o hacer, deberá revisarse antes de utilizarse con información institucional sensible.

4. Gestión de usuarios

El acceso al entorno institucional de inteligencia artificial deberá encontrarse vinculado a las funciones que cada persona desempeñe.

No todas las personas necesitan acceder a todos los proyectos, repositorios o capacidades.

La incorporación de un usuario deberá permitir determinar qué herramientas necesita y qué nivel de información podrá consultar.

En proyectos especialmente sensibles podrá crearse acceso restringido únicamente para quienes participen directamente.

Incorporación

Al ingresar al sistema institucional, la persona deberá:

1. recibir una cuenta o acceso institucional individual;
2. activar los mecanismos de seguridad exigidos;
3. conocer las reglas esenciales de este Manual;
4. aceptar la Declaración de Uso Responsable;
5. recibir los permisos correspondientes a sus funciones; y
6. completar la capacitación inicial cuando corresponda.

Cambio de funciones

Cuando una persona cambie de área o responsabilidades, deberán revisarse sus accesos.

No deberá mantenerse indefinidamente acceso a investigaciones o repositorios únicamente porque en algún momento la persona participó en ellos.

Salida

Cuando termine la relación de una persona con Fundamedios, deberán retirarse oportunamente sus accesos y transferirse la información institucional necesaria para garantizar continuidad.

La salida deberá comprender también asistentes personalizados, proyectos, automatizaciones, integraciones u otros recursos de IA que hubieran quedado bajo su administración.

5. Capacitación

La autorización para utilizar herramientas institucionales deberá acompañarse de capacitación suficiente para comprender sus riesgos y responsabilidades.

La capacitación no deberá limitarse a enseñar “cómo utilizar” una plataforma.

Una persona puede saber formular excelentes prompts y, al mismo tiempo, desconocer qué información no debe proporcionar al sistema.

La formación deberá cubrir progresivamente cuatro capacidades:

utilizar la herramienta; proteger la información; verificar sus resultados; reconocer cuándo no debe confiarse en ellos.

Los contenidos podrán adaptarse a las funciones.

Un investigador podrá necesitar mayor formación en búsqueda, verificación y análisis documental. Un editor podrá requerir mayor énfasis en detección de alucinaciones, atribuciones y control de borradores. Quien administre sistemas necesitará conocimientos adicionales sobre permisos, integraciones y seguridad.

5.1. Capacitación mediante casos reales

Siempre que resulte posible, las capacitaciones deberán incorporar ejemplos derivados de situaciones reales o plausibles para Fundamedios.

Será más útil mostrar cómo una IA puede inventar una sentencia, eliminar una atribución durante un resumen o confundir una fecha dentro de una investigación que limitarse a explicar abstractamente que “la IA puede cometer errores”.

Los incidentes y errores detectados durante la aplicación del Manual podrán convertirse, una vez protegida la información correspondiente, en casos de aprendizaje institucional.

6. Evaluación periódica de riesgos

La gestión de riesgos no deberá realizarse una sola vez al adoptar este Manual.

Las herramientas, capacidades y formas de utilización cambiarán.

Fundamedios deberá revisar periódicamente los riesgos asociados con su ecosistema de inteligencia artificial y también hacerlo cuando ocurra un cambio significativo.

La evaluación deberá preguntarse, fundamentalmente:

¿Qué estamos haciendo ahora con inteligencia artificial que no hacíamos cuando realizamos la última evaluación y qué riesgos nuevos produce esa capacidad?

Podrá revisarse, entre otros aspectos:

- nuevas herramientas;
- nuevas integraciones;
- información a la que tienen acceso;
- cambios en las políticas de los proveedores;
- incidentes ocurridos;
- errores recurrentes;
- nuevas obligaciones jurídicas;
- utilización de IA generativa audiovisual;
- automatización de procesos;
- funcionamiento de los controles humanos; y
- adecuación de las medidas de seguridad.

6.1. Matriz de riesgos

Para los usos más relevantes podrá mantenerse una matriz sencilla que permita priorizar controles.

No será necesario asignar puntuaciones matemáticas a cada actividad. Podrá utilizarse una clasificación cualitativa:

Riesgo	Ejemplo	Tratamiento
Bajo	Corrección de estilo de información pública	Control ordinario

Riesgo	Ejemplo	Tratamiento
Medio	Resumen de documentos para una nota	Verificación contra originales
Alto	Análisis de investigación no publicada	Entorno restringido + trazabilidad + revisión reforzada
Crítico	Información que pueda identificar una fuente protegida	Acceso excepcional y controles específicos

La clasificación deberá considerar conjuntamente la **probabilidad de error o exposición y la gravedad de sus consecuencias**.

7. Excepciones

Un Manual operativo debe poder responder a situaciones excepcionales.

Puede ocurrir que una herramienta no autorizada contenga una funcionalidad indispensable para una investigación concreta o que exista una necesidad técnica que no hubiera sido prevista.

La excepción no deberá convertirse en una vía informal para evitar las reglas.

Cuando exista una necesidad justificada, deberá evaluarse previamente:

- **la finalidad concreta;**
- **la información que se utilizará;**
- **los riesgos;**
- **las medidas de mitigación;**
- **el tiempo durante el cual se autorizará;** y
- **qué ocurrirá con la información después de terminar el uso excepcional.**

Las excepciones que involucren información de alto riesgo deberán quedar documentadas.

Una autorización excepcional para un caso determinado no incorporará automáticamente la herramienta al catálogo general.

8. Incumplimientos

El incumplimiento de este Manual deberá analizarse conforme a su naturaleza, gravedad, consecuencias y circunstancias.

No todo incumplimiento tendrá la misma entidad.

Deberá diferenciarse entre un error razonable cometido dentro del procedimiento, un desconocimiento que revele una necesidad de capacitación, una actuación negligente, una utilización reiteradamente contraria a las instrucciones y una evasión deliberada de los controles.

La respuesta institucional podrá incluir corrección del procedimiento, capacitación adicional, modificación de permisos, restricción temporal o permanente de determinadas herramientas y las demás medidas que correspondan conforme a las reglas internas y relaciones jurídicas aplicables.

Especial consideración merecerán las actuaciones deliberadas destinadas a:

extraer información institucional hacia cuentas personales; compartir información confidencial sin autorización; eludir controles de acceso; publicar contenido sin la revisión exigida; ocultar incidentes; alterar registros; o utilizar sistemas de IA para finalidades incompatibles con las funciones institucionales.

La determinación de consecuencias deberá realizarse conforme a las normas aplicables y no exclusivamente con base en este Manual.

9. Revisión del Manual

Este Manual deberá considerarse un instrumento sujeto a actualización.

La velocidad con la que evolucionan los sistemas de inteligencia artificial hace inadecuado asumir que las reglas tecnológicas específicas permanecerán vigentes indefinidamente.

La revisión podrá producirse:

periódicamente; ante modificaciones relevantes del marco jurídico; después de un incidente significativo; cuando se incorporen tecnologías sustancialmente diferentes; o cuando la experiencia demuestre que una regla resulta insuficiente o impracticable.

No toda actualización tecnológica requerirá modificar el Manual.

Precisamente por esta razón, sus reglas deberán concentrarse en principios, procedimientos y controles suficientemente estables, mientras que el catálogo de herramientas, configuraciones técnicas y determinadas instrucciones operativas podrán mantenerse en documentos complementarios de actualización más frecuente.

10. Documentos complementarios

Para evitar que el Manual se vuelva excesivamente extenso o quede rápidamente desactualizado, Fundamedios podrá mantener determinados instrumentos operativos separados.

Particularmente:

A. Catálogo de herramientas autorizadas.

B. Matriz institucional de riesgos de IA.

C. Plantillas de prompts aprobadas para actividades recurrentes.

D. Lista de comprobación previa a publicación.

E. Formato de registro de incidentes.

F. Declaración de Uso Responsable de Inteligencia Artificial.

G. Procedimiento técnico de incorporación y retiro de usuarios.

H. Protocolo de respuesta frente a incidentes de IA y seguridad.

Estos instrumentos formarán parte del sistema de gobernanza, pero podrán actualizarse sin necesidad de reescribir el contenido completo del Manual cada vez que cambie una herramienta o procedimiento técnico.

11. Responsabilidad institucional

La gobernanza de inteligencia artificial deberá partir de una responsabilidad compartida, pero no indeterminada.

Fundamedios deberá proporcionar herramientas adecuadas, establecer reglas comprensibles, administrar los accesos, capacitar a sus usuarios y mantener controles razonables.

Las personas usuarias deberán respetar esos controles, utilizar los entornos autorizados, proteger la información, verificar los resultados y comunicar los incidentes.

Los responsables editoriales deberán ejercer la supervisión correspondiente.

Los responsables tecnológicos deberán mantener condiciones adecuadas de seguridad y administración.

Y quienes tengan competencias jurídicas, de protección de datos o especializadas deberán intervenir cuando la naturaleza del riesgo lo requiera.

La existencia de responsabilidades distribuidas no significa que todos respondan por todo. El sistema deberá permitir identificar **qué función correspondía a cada participante y qué actuación realizó efectivamente**.

12. Principio de gobernanza

La inteligencia artificial institucional deberá permanecer bajo control de Fundamedios y no de las cuentas individuales de quienes temporalmente trabajan con ella.

Esto comprende las herramientas, pero también algo progresivamente más importante: **el conocimiento que la organización acumula mediante su utilización**.

Las metodologías desarrolladas, instrucciones perfeccionadas, bases de conocimiento autorizadas, estructuras editoriales, aprendizajes derivados de errores y demás elementos producidos dentro del trabajo institucional deberán permanecer disponibles para la organización conforme a las reglas de conservación aplicables.

Al mismo tiempo, esta memoria institucional deberá estar gobernada: no todo deberá conservarse, no todos deberán acceder a todo y no toda información conservada deberá estar disponible permanentemente para una IA.

La regla de cierre del capítulo será:

Fundamedios deberá mantener control institucional sobre las herramientas de inteligencia artificial que utiliza, las personas que acceden a ellas, la información que pueden consultar y el conocimiento que se genera mediante su uso, garantizando que la tecnología permanezca subordinada a las finalidades, responsabilidades y controles de la organización.

CAPÍTULO

VI

Ciberseguridad, gestión de incidentes y continuidad



La utilización institucional de inteligencia artificial amplía las capacidades de Fundamedios para investigar, analizar y producir información, pero también crea nuevos puntos de acceso, procesamiento y almacenamiento que deben ser protegidos. Una herramienta de IA puede recibir documentos, conectarse con repositorios, mantener historiales, utilizar memorias, acceder a servicios externos o integrarse con otras aplicaciones. Cada una de estas capacidades puede representar una ventaja operativa y, al mismo tiempo, introducir un riesgo de seguridad.

La ciberseguridad aplicada a la inteligencia artificial no deberá entenderse únicamente como protección frente a ataques externos. Comprende también la prevención de exposiciones accidentales, utilización indebida de cuentas, permisos excesivos, instrucciones maliciosas contenidas en documentos, conexiones inseguras, incorporación de información a herramientas no autorizadas y cualquier otra circunstancia que pueda comprometer la confidencialidad, integridad o disponibilidad de la información institucional.

Fundamedios deberá adoptar medidas proporcionales al riesgo de la información y de las funciones que permita realizar a cada sistema. El objetivo no será eliminar toda posibilidad de incidente —lo que resultaría imposible—, sino reducir razonablemente su probabilidad, limitar sus consecuencias y disponer de mecanismos para detectarlo, contenerlo y responder adecuadamente.

La seguridad deberá acompañar al ciclo completo de utilización de inteligencia artificial:

> **ACCESO → USO → INTERCAMBIO DE INFORMACIÓN → ALMACENAMIENTO → INTEGRACIONES → DETECCIÓN → RESPUESTA → RECUPERACIÓN → APRENDIZAJE**

1. Seguridad desde el diseño

La seguridad deberá considerarse antes de incorporar una herramienta de inteligencia artificial al trabajo institucional y no solamente después de que ocurra un incidente.

Cuando se configure una nueva plataforma, proyecto, asistente, repositorio o integración, deberán evaluarse desde el inicio los accesos que realmente necesita, la información que podrá procesar, las conexiones que podrá establecer y las acciones que podrá ejecutar.

Como principio general deberá utilizarse **la configuración menos permisiva que permita realizar adecuadamente la función prevista**.

No deberá otorgarse acceso general a información, aplicaciones o repositorios simplemente porque la plataforma técnicamente permita conectarlos.

Una herramienta destinada a apoyar la corrección editorial de contenidos públicos no necesita, por esa sola finalidad, acceso al correo institucional, archivos internos, investigaciones en curso o bases que contengan información confidencial.

La seguridad deberá responder a la función.

1.1. Privilegios mínimos

Las personas y herramientas deberán disponer únicamente de los permisos necesarios para realizar las actividades que les hayan sido asignadas.

Este principio deberá aplicarse tanto a usuarios humanos como a sistemas automatizados.

Una integración de IA que únicamente necesita consultar determinados documentos no deberá recibir capacidad para eliminarlos o modificarlos si esa función no resulta indispensable.

Una persona que trabaja en un proyecto determinado no deberá recibir acceso automático a todas las investigaciones de Fundamedios.

Un asistente configurado para una actividad específica no deberá acceder a repositorios ajenos a esa actividad únicamente para ampliar su contexto.

Los permisos deberán poder ampliarse cuando exista una necesidad justificada y reducirse cuando esa necesidad desaparezca.

1.2. Separación de entornos

Cuando la naturaleza de la información lo justifique, Fundamedios deberá separar espacios de trabajo.

No toda información institucional deberá encontrarse disponible dentro de un único entorno general de IA.

Podrán existir, según las capacidades tecnológicas utilizadas:

un entorno general, para información institucional de utilización ordinaria;

espacios por proyecto, para investigaciones o actividades determinadas;

espacios restringidos, para información de mayor sensibilidad;

entornos de prueba, para experimentar con nuevas herramientas sin utilizar información real protegida; y

repositorios de archivo, que no sean consultados automáticamente por los sistemas activos.

Esta separación permitirá limitar el impacto de un error, acceso indebido o funcionamiento inesperado de una herramienta.

---

2. Seguridad de cuentas y autenticación

Las cuentas institucionales de inteligencia artificial deberán protegerse conforme a estándares adecuados de seguridad.

Cada persona deberá utilizar, cuando la plataforma lo permita, una identidad individual dentro del entorno institucional. Deberá evitarse el uso ordinario de una única cuenta y contraseña compartidas entre múltiples usuarios.

Las credenciales deberán ser personales e intransferibles.

No deberán enviarse contraseñas mediante mensajes, incorporarse a documentos de trabajo, introducirse en prompts ni almacenarse en espacios accesibles sin protección.

Cuando la plataforma lo permita, deberá habilitarse autenticación multifactor para las cuentas institucionales y, especialmente, para aquellas que tengan privilegios administrativos.

2.1. Cuentas administrativas

Las cuentas con capacidad para incorporar usuarios, modificar configuraciones, habilitar integraciones, cambiar políticas de seguridad o acceder a funciones equivalentes deberán recibir protección reforzada.

El número de personas con privilegios administrativos deberá limitarse.

Las actividades ordinarias no deberán realizarse desde una cuenta administrativa cuando pueda utilizarse una cuenta con permisos normales.

Cuando una persona deje de necesitar privilegios elevados, estos deberán retirarse aunque mantenga su relación con la organización.

2.2. Sesiones y dispositivos

Los usuarios deberán procurar que las sesiones institucionales de IA sean utilizadas únicamente desde dispositivos razonablemente seguros.

Deberá evitarse mantener sesiones institucionales abiertas en dispositivos públicos o compartidos sin los controles correspondientes.

Cuando exista pérdida, robo o compromiso de un dispositivo que mantenga acceso a sistemas institucionales, deberá comunicarse oportunamente para evaluar el cierre de sesiones, cambio de credenciales o adopción de otras medidas.

---

3. Protección de la información durante su utilización

La clasificación de información desarrollada en los capítulos anteriores deberá determinar qué puede introducirse en cada herramienta.

El hecho de que una persona tenga acceso legítimo a determinada información **no significa automáticamente que pueda proporcionarla a cualquier sistema de inteligencia artificial**.

Antes de cargar información deberá considerarse:

- **qué información contiene;**
- **si la herramienta se encuentra autorizada para recibirla;**
- **si existe una versión menos sensible que permita realizar la misma tarea;**
- **si realmente es necesario proporcionar el documento completo;** y
- **qué ocurrirá con la información después de ser procesada.**

Cuando sea posible trabajar con información anonimizada, seudonimizada, fragmentada o depurada sin afectar la finalidad de la tarea, deberá preferirse esa alternativa.

3.1. Información que requiere protección reforzada

Deberá aplicarse especial cautela cuando el material contenga:

- identidad o información capaz de revelar una fuente confidencial;
- datos personales sensibles o especialmente protegidos;
- información sobre víctimas o personas en situación de riesgo;
- investigaciones no publicadas;
- estrategias jurídicas;
- credenciales o mecanismos de autenticación;
- información de seguridad;
- documentos sujetos a obligaciones de confidencialidad;
- información cuya exposición pueda comprometer a terceros; o
- cualquier material clasificado internamente con un nivel equivalente de restricción.

En determinados casos, el tratamiento adecuado no será utilizar controles adicionales, sino **no introducir la información en una herramienta generativa**.

4. Carga segura de archivos y documentos

Los archivos proporcionados a sistemas de IA deberán recibir el mismo tratamiento de seguridad que cualquier otra transferencia de información institucional.

Antes de cargar un archivo deberá verificarse que corresponda efectivamente al documento que se pretende utilizar y, cuando resulte necesario, que no incluya información adicional innecesaria.

Un documento puede contener más información de la que aparece inmediatamente en pantalla.

Deberá prestarse atención, según el formato, a:

metadatos; comentarios; control de cambios; hojas ocultas; columnas ocultas; archivos adjuntos; información incrustada; nombres de autores; datos de ubicación; versiones anteriores; y otros elementos que puedan permanecer incorporados al archivo.

Cuando resulte apropiado, deberá utilizarse una copia preparada específicamente para procesamiento mediante IA en lugar del archivo maestro.

4.1. Archivos de procedencia externa

Los documentos obtenidos de terceros, internet, correos electrónicos u otras fuentes externas deberán considerarse potencialmente no confiables desde una perspectiva tecnológica.

No deberá asumirse que un archivo es seguro únicamente porque su contenido aparente ser legítimo.

Cuando corresponda, deberá ser sometido a los controles de seguridad establecidos por la organización antes de abrirlo, ejecutarlo o incorporarlo a sistemas internos.

Esta precaución será especialmente importante respecto de archivos ejecutables, documentos con componentes activos, archivos comprimidos, enlaces incorporados o formatos capaces de ejecutar instrucciones.

5. Instrucciones maliciosas y prompt injection

Los sistemas de inteligencia artificial capaces de analizar documentos, navegar por internet o consultar repositorios pueden encontrar instrucciones incorporadas dentro del propio contenido que procesan.

Estas instrucciones pueden intentar modificar el comportamiento de la herramienta.

Un documento o página web podría incluir, por ejemplo, instrucciones destinadas a que el sistema ignore las órdenes del usuario, revele información disponible en su contexto, consulte otros archivos, utilice determinadas fuentes o ejecute una acción no solicitada.

Este riesgo se conoce generalmente como **prompt injection**.

El contenido analizado deberá tratarse como **información**, no como autoridad para modificar las instrucciones institucionales.

Una instrucción encontrada dentro de una página web, PDF, correo electrónico, documento, base de datos u otra fuente no deberá tener capacidad por sí sola para alterar la finalidad de la tarea.

5.1. Señales de posible manipulación

Los usuarios deberán considerar especialmente sospechosos aquellos resultados en los que una herramienta:

- solicite información que no necesita para realizar la tarea;

- intente acceder a archivos no relacionados;
- cambie repentinamente el objetivo de la consulta;
- solicite revelar instrucciones internas;
- proponga desactivar controles;
- intente ejecutar acciones no solicitadas;
- recomiende proporcionar credenciales;
- pretenda comunicarse con terceros sin que ello forme parte de la tarea; o
- produzca un comportamiento sustancialmente diferente después de analizar una fuente determinada.

Cuando exista una señal de este tipo, deberá detenerse el proceso y evaluarse el origen del comportamiento antes de continuar.

6. Integraciones con sistemas institucionales

La conexión de una herramienta de inteligencia artificial con correo electrónico, almacenamiento documental, bases de datos, sistemas editoriales, calendarios, redes sociales u otras plataformas aumenta significativamente sus capacidades y también el alcance potencial de un incidente.

Toda integración deberá responder a una finalidad institucional definida.

Antes de habilitarla deberá determinarse:

a) qué información podrá consultar; b) si podrá modificar información; c) si podrá realizar acciones externas; d) qué usuarios podrán utilizarla; e) qué registros genera; f) cómo puede revocarse; y **g) qué consecuencias tendría una utilización incorrecta.**

Deberá preferirse, cuando sea suficiente, el acceso de solo lectura frente al acceso de modificación.

6.1. Acciones externas

Deberá aplicarse un control especialmente estricto a sistemas capaces de realizar acciones en nombre de Fundamedios.

Esto comprende, por ejemplo:

enviar comunicaciones; publicar contenidos; modificar documentos; eliminar información; crear usuarios; realizar cambios en sistemas externos; o ejecutar cualquier otra actuación que produzca efectos fuera de la conversación con la herramienta.

La capacidad técnica para ejecutar una acción no deberá equivaler a autorización para decidir realizarla.

Cuando una acción pueda producir consecuencias relevantes, deberá existir confirmación o aprobación humana antes de su ejecución.

7. Protección frente a extracción de información

Los sistemas institucionales deberán configurarse, en la medida técnicamente posible, para reducir la posibilidad de que usuarios, herramientas o terceros obtengan información a la que no deberían acceder.

No deberá asumirse que una base de conocimiento conectada a una IA conserva automáticamente las mismas barreras de acceso que existían en los documentos originales.

Antes de conectar repositorios deberá comprobarse que la herramienta respete adecuadamente los permisos correspondientes.

Deberán realizarse pruebas razonables para determinar si un usuario puede obtener mediante preguntas indirectas información almacenada en espacios a los cuales no debería tener acceso.

Particular atención deberá prestarse a sistemas que utilicen múltiples repositorios o que generen respuestas combinando información proveniente de diferentes fuentes.

8. Credenciales, claves y secretos

No deberán proporcionarse a herramientas de IA contraseñas, claves privadas, códigos de recuperación, tokens de autenticación, secretos de aplicaciones u otras credenciales salvo que exista una integración institucional específicamente diseñada y autorizada para manejar ese tipo de información.

Las credenciales no deberán utilizarse como parte ordinaria de un prompt.

Cuando una persona introduzca accidentalmente una credencial en un sistema de IA, deberá considerarse potencialmente comprometida y comunicarse el incidente para determinar si corresponde revocarla o reemplazarla.

No será suficiente eliminar posteriormente la conversación si existe posibilidad de que la credencial haya quedado registrada en otros sistemas.

9. Detección de incidentes

Un incidente relacionado con inteligencia artificial no requiere necesariamente que exista un ataque informático.

Se considerará incidente cualquier acontecimiento que comprometa o pueda comprometer significativamente la confidencialidad, integridad, disponibilidad, trazabilidad o utilización legítima de la información o de los sistemas institucionales.

Entre otros, podrán constituir incidentes:

a) carga accidental de información confidencial en una herramienta no autorizada; b) acceso de una persona a información que no debía conocer; c) pérdida de una cuenta o dispositivo con acceso institucional; d) compromiso de credenciales; e) publicación automatizada no autorizada; f) extracción indebida de información; g) alteración de datos; h) comportamiento anómalo de una integración; i) exposición de una fuente; j) generación o utilización de contenido falso que llegue a un producto

institucional; k) incorporación de información incorrecta a una base de conocimiento; l) ejecución de una acción externa no autorizada; o **m)cualquier situación que pueda producir un daño relevante aunque finalmente no**

lo haya producido.

También deberán registrarse, cuando resulten relevantes, los incidentes evitados o detectados antes de producir consecuencias, pues pueden revelar una vulnerabilidad que requiera corrección.

10. Comunicación interna de incidentes

Las personas que identifiquen un posible incidente deberán comunicarlo oportunamente por el canal institucional definido para ese efecto.

No deberán intentar ocultarlo ni retrasar su comunicación por haber sido originado en un error propio.

La comunicación temprana permite limitar las consecuencias.

El reporte inicial no necesitará contener una investigación completa. Deberá proporcionar, en la medida conocida:

qué ocurrió;

cuándo ocurrió;

qué herramienta estuvo involucrada;

qué información pudo resultar afectada;

qué usuarios o terceros podrían encontrarse involucrados;

qué acciones ya se realizaron; y

si el incidente continúa activo.

La prioridad inicial será permitir la contención.

11. Respuesta inmediata

Ante un incidente deberá actuarse de acuerdo con su naturaleza y gravedad.

Las primeras acciones podrán comprender:

1. detener la actividad comprometida;
2. suspender temporalmente una integración;
3. cerrar sesiones;
4. revocar credenciales;
5. restringir accesos;
6. impedir nuevas publicaciones;
7. retirar temporalmente información expuesta cuando corresponda;
8. preservar los registros necesarios para investigar lo ocurrido; y
9. comunicar el incidente a las personas responsables.

No todas estas medidas serán necesarias en todos los casos.

La respuesta deberá procurar evitar dos errores opuestos: reaccionar de manera desproporcionada ante incidentes menores o minimizar un incidente que continúa exponiendo información.

12. Preservación de evidencia

Cuando ocurra un incidente relevante, deberá conservarse la información necesaria para determinar qué sucedió.

Podrán ser necesarios:

registros de acceso; conversaciones relevantes; prompts; respuestas del sistema; archivos involucrados; versiones; comunicaciones; configuraciones; fechas y horas; usuarios; acciones realizadas; y cualquier otro elemento que permita reconstruir el acontecimiento.

La preservación deberá realizarse de forma que no amplíe innecesariamente la exposición de la información comprometida.

No deberá alterarse o eliminarse evidencia relevante únicamente para “limpiar” el sistema antes de comprender el alcance del incidente.

13. Evaluación del impacto

Una vez contenido el incidente deberá determinarse razonablemente su alcance.

La evaluación deberá considerar:

qué información fue afectada;

si efectivamente salió del entorno autorizado;

quién pudo acceder a ella;

durante cuánto tiempo;

si puede seguir siendo utilizada o recuperada;

si existe riesgo para una fuente o tercero;

si fueron afectados datos personales;

si se modificó información;

si algún contenido incorrecto llegó al público;

si existen obligaciones jurídicas de comunicación; y

qué consecuencias adicionales podrían producirse.

La clasificación inicial podrá modificarse a medida que aparezca nueva información.

14. Incidentes relacionados con datos personales

Cuando un incidente involucre datos personales, deberá activarse además el procedimiento institucional correspondiente en materia de protección de datos.

La respuesta deberá considerar las obligaciones previstas en la legislación ecuatoriana y en las políticas institucionales aplicables, incluyendo la evaluación de posibles vulneraciones de seguridad, las medidas de mitigación y las comunicaciones que jurídicamente correspondan.

El hecho de que la exposición se haya producido dentro de una herramienta de inteligencia artificial no modifica la naturaleza de los datos ni elimina las obligaciones asociadas a su tratamiento.

Los registros del incidente deberán permitir determinar qué información personal fue involucrada y qué acciones se adoptaron.

15. Incidentes que afecten a fuentes confidenciales

La posible identificación o exposición de una fuente deberá tratarse como un incidente de especial gravedad.

La prioridad será determinar si existe un riesgo real para la persona y limitar inmediatamente cualquier posibilidad adicional de acceso.

El análisis deberá considerar no solamente si se reveló directamente su nombre, sino si la combinación de información expuesta permite razonablemente inferir su identidad.

Cuando sea necesario adoptar medidas adicionales para proteger a la fuente, estas deberán priorizarse frente a consideraciones operativas ordinarias.

Los detalles del incidente deberán mantenerse bajo acceso restringido.

16. Incidentes relacionados con contenidos publicados

Cuando un fallo relacionado con IA haya producido información incorrecta que llegó al público, deberán activarse simultáneamente dos procesos:

el proceso editorial de corrección, destinado a reparar la información publicada; y

el proceso de incidente, destinado a determinar por qué los controles permitieron que el error llegara a publicación.

Corregir el contenido no cerrará automáticamente el incidente.

Deberá determinarse si el error provino de una alucinación, una fuente incorrecta, un prompt inadecuado, una falla de extracción, una modificación automática, una revisión insuficiente o cualquier otra causa.

La finalidad será evitar que el mismo mecanismo reproduzca el problema.

17. Comunicación externa de incidentes

No todos los incidentes requerirán comunicación pública.

La decisión deberá considerar las obligaciones jurídicas aplicables, los derechos de las personas afectadas, la naturaleza de la información, el alcance del incidente y la necesidad de mantener una relación transparente con el público.

Cuando exista una obligación de informar a una autoridad, titular de datos, fuente, contraparte u otro tercero, la comunicación deberá realizarse conforme al procedimiento y dentro de los plazos que correspondan.

Cuando se realice una comunicación pública, esta deberá ser precisa y evitar especulaciones sobre aspectos que todavía se encuentren bajo investigación.

18. Recuperación

La respuesta a un incidente no terminará con su contención.

Fundamedios deberá determinar cuándo y bajo qué condiciones puede restablecerse la herramienta, cuenta, integración o procedimiento afectado.

Antes de restablecerlo deberá comprobarse, según corresponda:

- que las credenciales comprometidas hayan sido reemplazadas;
- que los permisos incorrectos hayan sido corregidos;
- que la vulnerabilidad identificada haya sido mitigada;
- que la información incorrecta haya sido retirada de repositorios activos;
- que las configuraciones sean adecuadas;

- que los usuarios afectados conozcan los cambios; y
- que exista un nivel razonable de confianza para reanudar la operación.

Cuando no pueda determinarse adecuadamente la causa o el riesgo permanezca elevado, podrá mantenerse suspendida una funcionalidad aunque ello produzca inconvenientes operativos.

19. Continuidad de las actividades

Fundamedios no deberá estructurar procesos críticos de manera que dependan absolutamente de la disponibilidad de una única herramienta de inteligencia artificial.

Los servicios pueden experimentar interrupciones, modificar funcionalidades, restringir cuentas, cambiar condiciones o dejar de encontrarse disponibles.

Para actividades relevantes deberá existir capacidad razonable de continuar el trabajo mediante procedimientos alternativos.

Esto podrá requerir:

conservar acceso a las fuentes originales;

mantener copias institucionales de documentos relevantes;

no depender exclusivamente de la memoria de una plataforma;

preservar las metodologías y prompts institucionales fuera de una única herramienta cuando resulte necesario;

conocer procedimientos manuales esenciales; y

establecer alternativas para actividades críticas.

La inteligencia artificial deberá aumentar la capacidad institucional, no convertirse en un punto único de falla.

20. Copias de respaldo y recuperación de información

La información institucional necesaria para continuidad, defensa jurídica, archivo o cumplimiento no deberá conservarse exclusivamente dentro de conversaciones con una herramienta de IA.

Cuando un resultado sea relevante deberá trasladarse o conservarse dentro de los sistemas institucionales apropiados conforme a las reglas de archivo del Manual.

Las copias de respaldo deberán protegerse conforme al nivel de sensibilidad de la información que contienen.

La existencia de un respaldo no justificará conservar indefinidamente información que debería eliminarse conforme a los períodos de conservación aplicables.

21. Dependencia de proveedores

La utilización de servicios externos implica una relación de dependencia tecnológica que deberá ser administrada.

Fundamedios deberá evitar, cuando resulte razonablemente posible, que conocimiento esencial para sus operaciones exista únicamente en un formato imposible de recuperar o trasladar.

Deberá conocerse qué información puede exportarse, qué ocurre cuando termina una suscripción y cómo se recuperan los materiales institucionales necesarios.

Las plantillas, metodologías, políticas y otros elementos propios de Fundamedios deberán conservarse de manera que la organización mantenga control sobre ellos independientemente del proveedor utilizado.

22. Aprendizaje posterior al incidente

Todo incidente significativo deberá producir una revisión posterior.

La pregunta no deberá limitarse a **"quién cometió el error"**, sino comprender **"qué condiciones permitieron que ocurriera y qué debe cambiar para reducir la posibilidad de repetición"**.

El análisis podrá identificar fallas en:

- **capacitación;**
- **permisos;**
- **configuración;**
- **selección de herramientas;**
- **clasificación de información;**
- **prompts;**
- **supervisión;**
- **procedimientos editoriales;**
- **integraciones;**
- **comunicación interna;** o
- **el propio Manual.**

Las medidas correctivas deberán dirigirse a la causa correspondiente.

Si una regla resulta sistemáticamente difícil de cumplir, deberá analizarse también si el procedimiento necesita ser rediseñado en lugar de atribuir automáticamente todos los incumplimientos a los usuarios.

23. Pruebas y revisión periódica de seguridad

Los controles establecidos deberán revisarse periódicamente para comprobar que continúan funcionando.

Podrán realizarse ejercicios de revisión sobre:

- usuarios con acceso;
- privilegios administrativos;
- integraciones activas;
- repositorios conectados;
- herramientas que ya no se utilizan;
- sesiones o dispositivos autorizados;
- información almacenada en memorias activas;
- mecanismos de autenticación;
- capacidad de retirar accesos; y
- procedimientos de respuesta ante incidentes.

En sistemas que manejen información de mayor riesgo podrán realizarse pruebas adicionales para comprobar si los controles de acceso y separación funcionan como se espera.

La finalidad no será únicamente encontrar vulnerabilidades técnicas, sino detectar **configuraciones, hábitos y procesos que produzcan exposición innecesaria**.

24. Seguridad y facilidad de uso

Los controles de seguridad deberán diseñarse considerando las condiciones reales de trabajo.

Cuando una medida sea innecesariamente compleja, lenta o incompatible con las necesidades operativas, aumenta el riesgo de que las personas busquen mecanismos para evitarla.

Esto no significa reducir la protección para hacerla cómoda.

Significa procurar que el procedimiento seguro sea también el procedimiento ordinario y accesible.

La disponibilidad de una cuenta institucional adecuada, herramientas autorizadas suficientemente capaces, instrucciones comprensibles y mecanismos rápidos para solicitar excepciones constituye también una medida de seguridad.

25. Regla de cierre de ciberseguridad y continuidad

La seguridad del sistema de inteligencia artificial dependerá tanto de las características técnicas de las herramientas como de la forma en que Fundamedios configure, utilice y supervise esas herramientas.

Ningún control tecnológico sustituirá completamente el criterio de las personas y ninguna instrucción interna será suficiente si las plataformas se encuentran configuradas de manera insegura.

Por ello, deberán operar conjuntamente **tecnología, procedimientos y supervisión humana**.

La regla general de este capítulo será:

La información de Fundamedios deberá permanecer protegida durante todo su ciclo de utilización con inteligencia artificial. Las herramientas recibirán únicamente los accesos y datos necesarios para cumplir su función; los incidentes deberán detectarse y comunicarse oportunamente; y la organización mantendrá capacidad para contenerlos, reconstruir lo ocurrido, recuperar sus operaciones y corregir las condiciones que los hicieron posibles.

CAPÍTULO

VII

Cumplimiento, responsabilidades y disposiciones finales



Este Manual establece el marco institucional para la utilización responsable de sistemas de inteligencia artificial en las actividades desarrolladas por Fundamedios y por sus iniciativas, proyectos, plataformas y denominaciones asociadas. Sus disposiciones deberán aplicarse conjuntamente con las demás políticas, procedimientos y obligaciones jurídicas que resulten aplicables a la organización.

La incorporación de inteligencia artificial a los procesos institucionales no modifica los deberes profesionales, editoriales, contractuales o legales que corresponden a las personas que intervienen en ellos. La tecnología constituye una herramienta de apoyo y, por tanto, su utilización deberá permanecer integrada dentro de las estructuras ordinarias de responsabilidad, supervisión y toma de decisiones de Fundamedios.

Al mismo tiempo, la existencia de un error relacionado con inteligencia artificial no permitirá atribuir automáticamente responsabilidad a la persona que operó la herramienta. Cuando ocurra un incidente o incumplimiento relevante deberán analizarse las circunstancias concretas, el procedimiento aplicable, las instrucciones existentes, los controles institucionales y la actuación de quienes participaron en las diferentes etapas.

El propósito de este capítulo es establecer las reglas finales que permitan interpretar, aplicar y exigir el cumplimiento del Manual de manera coherente.

1. Obligatoriedad del Manual

Las disposiciones de este Manual serán de aplicación obligatoria para las personas que utilicen herramientas de inteligencia artificial en actividades institucionales comprendidas dentro de su ámbito.

Su aplicación alcanzará a quienes intervengan en procesos de investigación, levantamiento de información, análisis, almacenamiento, producción editorial, verificación, comunicación, administración tecnológica, publicación o cualquier otra actividad en la que se utilicen sistemas de inteligencia artificial por cuenta o en beneficio de Fundamedios.

La modalidad de relación con la organización no deberá utilizarse para eludir los estándares de protección de información.

Cuando consultores, proveedores, investigadores externos u otros terceros deban acceder a información institucional o utilizar sistemas de IA dentro de un proyecto de Fundamedios, las obligaciones aplicables deberán incorporarse, según corresponda, a sus instrucciones, contratos, acuerdos de confidencialidad u otros instrumentos pertinentes.

2. Aplicación a las iniciativas y proyectos vinculados con Fundamedios

Las reglas establecidas en este Manual deberán aplicarse a los productos y actividades desarrollados directamente bajo la denominación Fundamedios y también a aquellos realizados a través de Ecuador Chequea, Liberties 360 y las demás iniciativas, proyectos, plataformas o denominaciones institucionales que formen parte de la organización.

La existencia de metodologías particulares para determinadas iniciativas no excluirá la aplicación del Manual.

Cuando una actividad cuente con reglas específicas —por ejemplo, una metodología propia de verificación, investigación o producción— estas deberán aplicarse conjuntamente con las disposiciones generales sobre inteligencia artificial.

Si existiera una incompatibilidad relevante entre una metodología específica y este Manual, deberá ser identificada y resuelta institucionalmente antes de utilizar IA en el proceso correspondiente.

3. Integración con otras políticas institucionales

Este Manual no sustituye las demás políticas de Fundamedios.

Deberá interpretarse conjuntamente con las normas internas que regulen, entre otras materias:

protección de datos personales;

seguridad de la información;

confidencialidad;

protección de fuentes;

metodologías periodísticas y de verificación;

archivo y conservación documental;

uso de sistemas tecnológicos;

propiedad intelectual;

relaciones laborales y contractuales; y

cualquier otro procedimiento relacionado con la actividad institucional.

Cuando una misma situación se encuentre sometida a diferentes instrumentos, deberán cumplirse conjuntamente las obligaciones aplicables.

La utilización de inteligencia artificial no constituirá una excepción frente a una prohibición o restricción establecida en otra política institucional.

4. Cumplimiento de la normativa aplicable

El uso institucional de inteligencia artificial deberá realizarse conforme al ordenamiento jurídico aplicable.

Especial atención deberá prestarse a las obligaciones relacionadas con protección de datos personales, derechos de las personas, confidencialidad, propiedad intelectual, libertad de expresión, responsabilidades derivadas de publicaciones, seguridad de la información y demás materias vinculadas con las actividades desarrolladas por Fundamedios.

Cuando una herramienta permita técnicamente realizar una actividad que se encuentre jurídicamente restringida, **prevalecerá la obligación jurídica**.

La capacidad tecnológica no constituye autorización legal.

Cuando exista una duda razonable sobre la licitud de un tratamiento de información o de una utilización específica de IA, deberá procurarse el criterio correspondiente antes de continuar cuando la naturaleza del riesgo así lo justifique.

5. Responsabilidad por el uso de inteligencia artificial

La utilización de una herramienta de IA no traslada a esa herramienta la responsabilidad por las decisiones adoptadas dentro de la organización.

Las personas deberán responder por las funciones que les hayan sido asignadas dentro del procedimiento y por el cumplimiento de los deberes asociados a esas funciones.

Sin embargo, la determinación de responsabilidad deberá realizarse considerando **el proceso completo y no únicamente quién ejecutó materialmente una instrucción en la herramienta**.

Un error puede originarse, entre otras causas, en:

- una fuente incorrecta;
- una clasificación inadecuada de información;
- instrucciones deficientes;
- una respuesta incorrecta del sistema;
- una herramienta inapropiada;
- una configuración institucional insuficiente;
- falta de capacitación;
- ausencia de controles;
- una revisión humana deficiente;
- una modificación posterior a la aprobación; o
- un incumplimiento deliberado de las reglas.

La identificación de la causa será necesaria para determinar las medidas que correspondan.

6. La supervisión humana no constituye una transferencia automática de responsabilidad

La existencia de una persona encargada de revisar o aprobar un producto no significa que todas las responsabilidades anteriores se transfieran automáticamente a ella.

Cada participante deberá cumplir adecuadamente la función que le corresponda.

Quien recopila información deberá aplicar los controles de recopilación.

Quien verifica deberá realizar las comprobaciones correspondientes.

Quien utiliza IA para producir un borrador deberá revisar los resultados de esa utilización.

Quien realiza una revisión especializada deberá responder por el alcance de esa revisión.

Quien autoriza una publicación deberá ejercer efectivamente la función de aprobación que le haya sido asignada.

La responsabilidad deberá analizarse conforme a las actuaciones y omisiones concretas de cada etapa.

7. Diferenciación entre error, negligencia e incumplimiento deliberado

La utilización de inteligencia artificial implica inevitablemente la posibilidad de errores. La existencia de un error no demostrará por sí sola que una persona haya incumplido este Manual.

Deberán diferenciarse al menos tres situaciones.

Error producido dentro del procedimiento

Existe cuando una persona aplicó razonablemente los controles establecidos y, pese a ello, se produjo un resultado incorrecto o un incidente que no era razonablemente detectable con los medios disponibles.

Estos casos deberán utilizarse principalmente para revisar y mejorar el sistema.

Incumplimiento por falta de diligencia

Existe cuando una persona omite controles que debía aplicar, utiliza una herramienta sin la revisión requerida, ignora advertencias relevantes o actúa sin el cuidado razonablemente exigible conforme a su función.

La respuesta deberá considerar la gravedad, reiteración y consecuencias.

Incumplimiento deliberado

Existe cuando una persona conoce una regla o restricción aplicable y decide conscientemente eludirla.

Podrán constituir ejemplos:

utilizar deliberadamente una cuenta personal para extraer información institucional restringida;

introducir información confidencial en una herramienta expresamente no autorizada;

publicar contenido sin la aprobación requerida;

eliminar registros para ocultar una actuación;

compartir credenciales;

desactivar controles de seguridad sin autorización; o

ocultar deliberadamente un incidente relevante.

Esta diferenciación deberá impedir tanto la ausencia de responsabilidad como la atribución automática de culpa por cualquier error tecnológico.

8. Determinación de responsabilidades ante un incidente

Cuando un incidente relevante requiera determinar responsabilidades, deberá reconstruirse el procedimiento antes de adoptar conclusiones.

El análisis podrá considerar:

I. qué tarea debía realizarse;

II. qué persona o personas tenían responsabilidad sobre ella;

III. qué herramienta fue utilizada;

IV. si se encontraba autorizada;

V. qué información recibió;

VI. qué instrucciones fueron proporcionadas;

VII. qué resultado produjo;

VIII. qué controles debían aplicarse;

IX. qué controles se aplicaron efectivamente;

X. qué advertencias o señales existieron;

XI. quién tomó las decisiones posteriores; y

XII. qué circunstancias institucionales pudieron contribuir al resultado.

La determinación no deberá basarse exclusivamente en el resultado final.

Un resultado incorrecto puede haberse producido pese a la aplicación adecuada del procedimiento, mientras que un resultado correcto puede haber sido obtenido mediante una utilización manifiestamente contraria a las reglas.

Ambas situaciones deberán distinguirse.

9. Responsabilidad institucional y responsabilidad individual

El sistema deberá reconocer que la utilización segura de IA depende simultáneamente de decisiones institucionales e individuales.

Corresponderá a Fundamedios, dentro del ámbito de sus posibilidades y obligaciones:

- proporcionar herramientas institucionales adecuadas;
- establecer reglas claras;
- definir accesos;
- mantener configuraciones razonables de seguridad;
- capacitar a las personas usuarias;
- establecer mecanismos de supervisión;
- mantener procedimientos de respuesta ante incidentes; y
- revisar periódicamente la eficacia del sistema.

Corresponderá a las personas usuarias:

- utilizar las herramientas conforme a las autorizaciones;
- proteger sus credenciales;
- respetar la clasificación de información;
- aplicar los controles establecidos;
- verificar los resultados cuando corresponda;
- solicitar orientación cuando exista una duda material; y
- comunicar oportunamente los incidentes.

La existencia de obligaciones institucionales no elimina las obligaciones individuales y viceversa.

10. Declaración de Uso Responsable

Toda persona que tenga acceso regular a los entornos institucionales de inteligencia artificial deberá conocer este Manual y aceptar la correspondiente **Declaración de Uso Responsable de Inteligencia Artificial**.

La declaración tendrá por finalidad dejar constancia de que la persona conoce las reglas esenciales que condicionan el acceso y utilización de las herramientas institucionales.

No deberá redactarse como una renuncia general de derechos ni como una transferencia automática de responsabilidad desde Fundamedios hacia el usuario.

Su finalidad será establecer compromisos concretos de actuación.

Entre ellos deberán encontrarse:

a) utilizar las herramientas institucionales autorizadas; b) proteger la información; c) no trasladar información institucional a cuentas personales sin autorización; d) respetar las restricciones aplicables a datos personales y fuentes; e) verificar los resultados cuando corresponda; f) mantener la supervisión humana; g) respetar los procedimientos de publicación; h) proteger las credenciales; i) comunicar incidentes; y **j) colaborar con los procedimientos de revisión cuando sea necesario.**

La declaración formará parte de los anexos operativos del Manual.

11. La Declaración de Uso Responsable no será una cláusula de exoneración

La aceptación de la Declaración de Uso Responsable no significará que Fundamedios quede automáticamente exonerado frente a cualquier consecuencia relacionada con la utilización de inteligencia artificial.

Tampoco significará que una persona usuaria asuma automáticamente responsabilidad individual por cualquier resultado incorrecto producido mediante una herramienta.

La responsabilidad frente a terceros se determinará conforme al ordenamiento jurídico aplicable.

Internamente, la declaración servirá para acreditar las reglas conocidas y aceptadas por el usuario y permitirá analizar si una determinada actuación se ajustó o no al procedimiento institucional.

Por tanto:

La declaración delimita obligaciones de uso; no sustituye el análisis jurídico de responsabilidad.

12. Prohibición de utilizar IA para eludir controles institucionales

Ninguna persona deberá utilizar inteligencia artificial con el propósito de evitar procedimientos que serían obligatorios si la misma actividad fuera realizada sin IA.

No podrá utilizarse una herramienta para:

- **evitar una revisión editorial;**
- **obtener indirectamente información a la que no se tiene acceso;**
- **eludir restricciones de una base de datos;**
- **reconstruir información deliberadamente anonimizada;**
- **evitar una revisión jurídica requerida;**
- **publicar mediante una automatización aquello que requería aprobación humana;**
- **transferir información hacia un sistema no autorizado;** o
- **realizar indirectamente cualquier actividad que se encuentre prohibida de manera directa.**

La utilización de una tecnología diferente no modifica la naturaleza de la actividad.

13. Propiedad y control de los productos institucionales

Los productos, metodologías, plantillas, prompts institucionales, configuraciones, bases de conocimiento y demás materiales desarrollados en el marco de las actividades de Fundamedios deberán administrarse conforme a las reglas jurídicas y contractuales aplicables.

La utilización de una cuenta institucional deberá procurar que estos elementos permanezcan bajo control de la organización y no exclusivamente dentro de espacios personales.

Cuando una persona desarrolle una metodología de trabajo mediante IA como parte de sus actividades institucionales, deberá procurarse que la organización pueda mantener acceso a los elementos necesarios para continuar utilizándola cuando corresponda.

Esta regla deberá aplicarse sin perjuicio de los derechos que legal o contractualmente correspondan a autores, trabajadores, contratistas, proveedores u otros terceros.

14. Confidencialidad posterior

La terminación de la relación de una persona con Fundamedios no convertirá en pública la información institucional a la que tuvo acceso mediante sistemas de IA.

Las obligaciones de confidencialidad que deban continuar después de terminada la relación se mantendrán conforme a las normas y acuerdos aplicables.

La persona no deberá conservar en cuentas personales copias de conversaciones, documentos, bases, prompts con información protegida u otros materiales institucionales cuya conservación no se encuentre autorizada.

Cuando corresponda, la salida deberá incluir la devolución, transferencia o eliminación de información conforme al procedimiento institucional.

15. Cooperación en investigaciones internas

Cuando exista un incidente razonablemente relacionado con el uso institucional de inteligencia artificial, las personas involucradas deberán colaborar con la reconstrucción del proceso conforme a sus obligaciones aplicables.

La investigación deberá procurar determinar hechos y causas antes de atribuir responsabilidades.

La colaboración podrá comprender la identificación de conversaciones relevantes, documentos, herramientas utilizadas, instrucciones proporcionadas y acciones realizadas.

El acceso a esta información deberá limitarse a aquello que resulte necesario para investigar el incidente y respetar las reglas de protección de datos, confidencialidad y demás derechos aplicables.

16. Proporcionalidad de las medidas

Las medidas adoptadas ante incumplimientos deberán guardar relación con la naturaleza de la conducta y el riesgo producido.

Según las circunstancias, podrá resultar suficiente:

- **una aclaración del procedimiento;**
- **capacitación adicional;**
- **modificación de una configuración;**
- **revisión de permisos;**
- **supervisión reforzada;**
- **restricción de determinadas funciones;** o
- **modificación de la metodología institucional.**

Cuando existan conductas graves o deliberadas podrán aplicarse las medidas previstas en las normas internas, contractuales o legales correspondientes.

El Manual no establecerá por sí mismo un régimen disciplinario autónomo cuando dicho régimen deba encontrarse contenido en otros instrumentos.

17. Protección de quien reporta un incidente de buena fe

La política institucional deberá favorecer la comunicación temprana de errores e incidentes.

Una persona que detecte que introdujo accidentalmente información en una herramienta incorrecta deberá tener incentivos para comunicarlo inmediatamente y permitir su contención, no para ocultarlo por temor a que el simple reporte sea interpretado como reconocimiento automático de responsabilidad.

La comunicación de buena fe no impedirá analizar la actuación cuando corresponda, pero deberá diferenciarse claramente de la ocultación deliberada de un incidente.

La cultura de cumplimiento deberá favorecer **detección temprana, transparencia interna y corrección**, sin eliminar la responsabilidad por actuaciones que efectivamente lo ameriten.

18. Interpretación del Manual

Las disposiciones deberán interpretarse conforme a la finalidad general del Manual: permitir que Fundamedios aproveche las capacidades de inteligencia artificial manteniendo control sobre la información, la calidad de sus productos, la seguridad, la protección de datos y las decisiones humanas relevantes.

Cuando una situación no se encuentre expresamente prevista, deberá atenderse a los principios desarrollados en el conjunto del documento.

No deberá interpretarse el silencio del Manual como autorización automática para cualquier utilización tecnológicamente posible.

Cuando exista una duda significativa deberá evaluarse el riesgo y, cuando corresponda, consultar a la persona responsable antes de continuar.

19. Actualización tecnológica sin reforma inmediata del Manual

La aparición de una nueva herramienta, modelo o funcionalidad no requerirá necesariamente reformar inmediatamente este Manual.

Los responsables institucionales podrán establecer instrucciones operativas temporales cuando resulte necesario responder a cambios tecnológicos, siempre que sean compatibles con los principios establecidos.

Las modificaciones permanentes de importancia deberán incorporarse posteriormente al Manual o a sus instrumentos complementarios.

Esto permitirá que Fundamedios responda con rapidez a cambios tecnológicos sin mantener un documento permanentemente desactualizado.

20. Revisión periódica

El Manual deberá revisarse periódicamente y, adicionalmente, cuando exista una circunstancia que justifique una evaluación extraordinaria.

Podrán justificarla:

- reformas legales relevantes;
- decisiones de autoridades competentes que afecten las actividades comprendidas;
- cambios sustanciales en las herramientas utilizadas;
- incorporación de agentes o automatizaciones con nuevas capacidades;
- incidentes significativos;
- resultados de auditorías;
- modificaciones importantes en los procesos de Fundamedios; o
- experiencia acumulada que demuestre la necesidad de cambiar una regla.

La revisión deberá evaluar no solamente si faltan nuevas reglas, sino también si existen disposiciones que han dejado de ser necesarias o que dificultan injustificadamente el trabajo.

21. Control de versiones del Manual

Cada versión aprobada del Manual deberá poder identificarse.

Como mínimo deberá constar:

- **número o identificación de versión;**
- **fecha de aprobación;**
- **fecha de entrada en vigencia;** y
- **cuando corresponda, fecha de la última actualización.**

Las modificaciones sustantivas deberán comunicarse a las personas sujetas al Manual.

Cuando una actualización altere obligaciones relevantes de los usuarios, deberá evaluarse la necesidad de capacitación adicional o de una nueva aceptación de la Declaración de Uso Responsable.

22. Instrumentos operativos y anexos

Los anexos y documentos operativos previstos en este Manual permitirán desarrollar sus reglas sin incorporar todos los detalles de aplicación dentro del cuerpo principal.

Podrán actualizarse con mayor frecuencia cuando ello resulte necesario por cambios tecnológicos u operativos.

Deberá procurarse que exista coherencia entre el Manual y estos instrumentos.

Cuando una plantilla, checklist o instrucción operativa contradiga una disposición del Manual, deberá prevalecer este último hasta que la contradicción sea revisada y corregida.

23. Entrada en vigencia

El Manual entrará en vigencia en la fecha que determine Fundamedios mediante el mecanismo institucional correspondiente.

A partir de su entrada en vigencia deberán implementarse progresivamente las medidas necesarias para su aplicación.

Cuando determinados controles requieran contratación, configuración tecnológica, migración de información, capacitación u otras acciones que no puedan ejecutarse inmediatamente, deberá establecerse un proceso razonable de implementación.

Durante ese período deberán priorizarse los controles relacionados con los riesgos de mayor gravedad, particularmente aquellos vinculados con información confidencial, datos personales, fuentes protegidas, seguridad de cuentas y publicación de contenidos.

24. Evaluación de implementación

Transcurrido un período razonable desde la entrada en vigencia, deberá realizarse una primera evaluación de implementación.

Esta revisión deberá analizar si:

las personas conocen el Manual;

existen herramientas institucionales suficientes;

las cuentas personales han dejado de utilizarse para actividades que deben permanecer dentro del entorno institucional;

los mecanismos de clasificación y almacenamiento funcionan;

los controles editoriales son aplicables en la práctica;

los incidentes pueden reportarse adecuadamente;

las responsabilidades se encuentran suficientemente claras; y

existen disposiciones que requieran ajustes.

Esta primera evaluación será particularmente importante porque permitirá contrastar el diseño del Manual con las condiciones reales de trabajo.

25. Disposición final

La incorporación de inteligencia artificial a las actividades de Fundamedios deberá producirse bajo un principio de **control institucional y responsabilidad humana**.

La herramienta podrá investigar, ordenar, comparar, resumir, analizar, traducir, asistir en la redacción, detectar inconsistencias y facilitar numerosos procesos. Ninguna de estas capacidades modificará el hecho de que las decisiones relevantes sobre la información, su interpretación y su publicación corresponden a personas.

Fundamedios deberá mantener control sobre **qué herramientas utiliza, qué información reciben, qué pueden hacer con ella, quién accede a los sistemas, cómo se verifican sus resultados y quién adopta las decisiones finales**.

Al mismo tiempo, deberá mantener capacidad para reconstruir el proceso cuando algo salga mal y distinguir entre un error tecnológico, una falla de procedimiento, una deficiencia institucional y un incumplimiento individual.

La regla que cierra este Manual será:

La inteligencia artificial deberá ampliar las capacidades de Fundamedios sin sustituir su criterio, comprometer la protección de su información ni diluir las responsabilidades humanas e institucionales que sustentan la confianza en su trabajo.

ANEXOS

1–9

Herramientas, matrices y formatos para aplicar el Manual



Anexo 1

Declaración de uso responsable de inteligencia artificial

La presente Declaración forma parte del **Manual para el Uso Responsable, Ético y Seguro de Inteligencia Artificial de Fundamedios** y establece los compromisos mínimos que deberán observar las personas autorizadas para utilizar herramientas, cuentas, perfiles, proyectos, asistentes, repositorios o entornos institucionales de inteligencia artificial.

Su aceptación tiene por objeto dejar constancia de que la persona usuaria conoce las condiciones bajo las cuales podrá utilizar inteligencia artificial en actividades vinculadas con Fundamedios, Ecuador Chequea, Liberties 360 y las demás iniciativas, programas, plataformas y proyectos administrados o desarrollados por la organización.

La Declaración no constituye una transferencia automática de responsabilidad hacia la persona usuaria ni una exoneración general de responsabilidad para Fundamedios. Su función es establecer obligaciones concretas de actuación y permitir que el uso institucional de inteligencia artificial pueda desarrollarse bajo criterios conocidos de seguridad, confidencialidad, verificación, trazabilidad y supervisión humana.

1. Conocimiento y aceptación del Manual

Declaro haber recibido acceso al Manual para el Uso Responsable, Ético y Seguro de Inteligencia Artificial de Fundamedios y conocer las reglas que resulten aplicables a las actividades que desempeño.

Me comprometo a utilizar las herramientas institucionales de inteligencia artificial de conformidad con dicho Manual, las políticas internas relacionadas y las instrucciones específicas que Fundamedios establezca para determinados proyectos, plataformas o tipos de información.

Cuando exista una duda material respecto de la aplicación de una regla, procuraré obtener la orientación correspondiente antes de realizar una actividad que pueda comprometer información, derechos de terceros, seguridad o productos institucionales.

2. Uso de cuentas y entornos institucionales

Me comprometo a utilizar, para las actividades institucionales, las cuentas, perfiles y entornos de inteligencia artificial autorizados por Fundamedios cuando estos se encuentren disponibles para la tarea correspondiente.

No utilizaré cuentas personales con el propósito de almacenar, procesar o desarrollar información institucional que deba permanecer dentro del entorno de Fundamedios.

En particular, me comprometo a no trasladar deliberadamente hacia cuentas personales:

- investigaciones no publicadas;
- documentos institucionales protegidos;
- bases de datos;
- información de fuentes;
- prompts que contengan información reservada;
- resultados que deban formar parte de un expediente institucional; o

- cualquier otro material cuya conservación corresponda a Fundamedios.

Tampoco utilizaré herramientas externas con el propósito de evitar restricciones, controles, límites de acceso o configuraciones de seguridad establecidos dentro del entorno institucional.

3. Protección de información institucional

Reconozco que proporcionar información a una herramienta de inteligencia artificial constituye una forma de procesamiento de esa información.

Antes de introducir un dato, documento, archivo o conjunto de información, verificaré que la herramienta utilizada se encuentre autorizada para recibirlo.

Me comprometo a aplicar los criterios de clasificación, minimización, confidencialidad y seguridad definidos por Fundamedios.

Cuando una tarea pueda realizarse utilizando menos información, retirando elementos personales o trabajando con una versión depurada del documento, procuraré utilizar esa alternativa.

4. Protección de datos personales

Me comprometo a tratar los datos personales utilizados mediante inteligencia artificial conforme a las reglas establecidas en el Manual y a las políticas institucionales aplicables.

No introduciré datos personales de manera indiscriminada por el solo hecho de encontrarse disponibles.

Antes de proporcionar información personal a una herramienta consideraré, cuando corresponda:

si es necesaria para la finalidad de la tarea;

si puede reducirse o anonimizarse;

si la herramienta está autorizada para procesarla; y

si su conservación o reutilización puede generar riesgos adicionales.

Prestaré especial atención cuando el material involucre datos sensibles, víctimas, menores de edad, denunciantes, personas amenazadas u otras categorías que requieran protección reforzada.

5. Protección de fuentes confidenciales

Reconozco que la identidad de una fuente periodística confidencial constituye información de protección reforzada.

No introduciré en herramientas de inteligencia artificial no autorizadas el nombre de una fuente confidencial ni información que permita razonablemente identificarla.

Cuando sea necesario procesar información proporcionada por una fuente protegida, procuraré separar su identidad del contenido que necesita ser analizado, utilizando los mecanismos institucionales previstos para ello.

No utilizaré inteligencia artificial para intentar reconstruir o descubrir la identidad de una fuente cuya confidencialidad Fundamedios se haya comprometido a proteger.

6. Información confidencial y restringida

No incorporaré información confidencial, reservada o restringida en herramientas que no hayan sido autorizadas para procesar ese nivel de información.

Esta obligación comprende tanto la introducción directa dentro de un prompt como la carga mediante documentos, imágenes, audios, enlaces privados, bases de datos, integraciones u otros mecanismos.

Cuando exista duda sobre el nivel de protección de un documento, evitaré introducirlo hasta determinar si el tratamiento se encuentra autorizado.

7. Protección de credenciales

No proporcionaré a herramientas de inteligencia artificial contraseñas, claves privadas, tokens, códigos de recuperación, credenciales de autenticación ni otros mecanismos que puedan permitir acceso a sistemas institucionales, salvo dentro de integraciones técnicas expresamente autorizadas para ello.

No compartiré mis credenciales institucionales con otras personas.

Cuando considere que una credencial pudo haber quedado expuesta, lo comunicaré oportunamente para que puedan adoptarse las medidas de seguridad correspondientes.

8. Verificación de la información

Reconozco que una respuesta producida por inteligencia artificial no constituye, por sí misma, una fuente ni una garantía de exactitud.

No utilizaré como hecho publicable una afirmación cuya única base sea una generación de IA.

Cuando la herramienta proporcione información factual relevante, procuraré identificar y revisar la fuente que la sustente de acuerdo con el nivel de riesgo correspondiente.

En particular, verificaré cuando corresponda:

nombres; fechas; cifras; citas; documentos; cargos; normas; sentencias; estudios; estadísticas; enlaces; antecedentes; y demás información cuya alteración pueda afectar el sentido o exactitud de un producto.

9. Prohibición de fabricar o completar información

No utilizaré inteligencia artificial para fabricar o completar como hechos datos que la investigación no haya podido establecer.

No presentaré como reales:

- citas generadas;
- documentos inexistentes;
- referencias fabricadas;
- hechos inferidos sin sustento;
- relaciones entre personas no verificadas;
- acusaciones construidas por la herramienta;
- imágenes o audios sintéticos presentados como evidencia auténtica; o

- cualquier otro contenido generado que pueda inducir a creer que existe evidencia cuando esta no ha sido obtenida.

Cuando una herramienta produzca información nueva que pueda ser relevante, la trataré como un hallazgo pendiente de verificación.

10. Supervisión humana

Me comprometo a mantener control humano sobre los resultados producidos mediante inteligencia artificial dentro de las actividades que se encuentren bajo mi responsabilidad.

No consideraré terminado un producto sustantivo únicamente porque la herramienta haya entregado una respuesta aparentemente completa.

Cuando me corresponda revisar un resultado, realizaré una revisión efectiva y no meramente formal.

Cuando una decisión exceda mis funciones, conocimientos o nivel de autorización, solicitaré la intervención correspondiente.

11. Producción de contenidos

Cuando utilice IA para elaborar borradores, resúmenes, adaptaciones u otros productos, procuraré que la herramienta trabaje con información previamente validada y con instrucciones suficientemente claras.

No permitiré deliberadamente que la IA:

elimine atribuciones necesarias;

transforme denuncias en hechos comprobados;

aumente el grado de certeza de las fuentes;

invente citas o cifras;

complete vacíos factuales;

elimine incertidumbres relevantes; o

incorpore información no verificada como parte del producto final.

Todo producto destinado a publicación seguirá los procedimientos de revisión y aprobación establecidos por Fundamedios.

12. Uso de imágenes, audio y video generados mediante IA

No utilizaré contenido sintético para hacer creer al público que una persona realizó una acción, pronunció una declaración o participó en un acontecimiento que no ocurrió de esa manera.

Cuando utilice material visual o audiovisual generado o modificado mediante inteligencia artificial, respetaré las reglas de identificación y transparencia previstas en el Manual.

No presentaré una imagen generada o modificada sustancialmente mediante IA como evidencia documental auténtica.

13. Trazabilidad

Cuando el nivel de riesgo o la naturaleza del trabajo así lo exija, conservaré los elementos necesarios para reconstruir razonablemente el uso de inteligencia artificial.

Esto podrá comprender fuentes, prompts relevantes, resultados sustantivos, verificaciones, versiones, registros de revisión y demás información prevista en el Manual.

No eliminaré deliberadamente registros cuya conservación sea necesaria para ocultar una actuación, un error o un incidente.

La trazabilidad deberá mantenerse dentro de los repositorios institucionales correspondientes y no depender exclusivamente de mis cuentas o dispositivos personales.

14. Protección del conocimiento institucional

Reconozco que las metodologías, prompts institucionales, configuraciones, estructuras de trabajo, bases de conocimiento y demás materiales generados dentro de actividades desarrolladas para Fundamedios pueden constituir conocimiento institucional cuya continuidad debe preservarse.

Me comprometo a mantener dentro del entorno institucional aquellos elementos que deban conservarse para permitir que Fundamedios continúe utilizando el conocimiento generado durante el trabajo.

No trasladaré sistemáticamente estos materiales a cuentas personales ni los mantendré exclusivamente bajo mi control cuando correspondan a procesos institucionales.

15. Seguridad de herramientas e integraciones

Utilizaré únicamente aquellas integraciones, conectores, extensiones o funcionalidades que hayan sido autorizadas para la información y finalidad correspondientes.

No concederé permisos adicionales a una aplicación únicamente porque estos sean solicitados durante su instalación.

Cuando una herramienta tenga capacidad para acceder a repositorios, correo, documentos u otros sistemas institucionales, respetaré los límites de acceso establecidos.

No utilizaré agentes o automatizaciones para ejecutar acciones externas que requieran autorización humana sin cumplir previamente el procedimiento correspondiente.

16. Incidentes y errores

Me comprometo a comunicar oportunamente cualquier error o incidente que pueda afectar:

información institucional; datos personales; fuentes confidenciales; seguridad de cuentas; productos publicados; reputación de terceros; integridad de sistemas; o cualquier otro interés protegido por el Manual.

La comunicación deberá realizarse aunque el incidente haya sido originado accidentalmente por una actuación propia.

Cuando detecte un problema, no destruiré evidencia relevante ni intentaré resolverlo de manera que impida posteriormente reconstruir lo ocurrido.

Adoptaré, dentro de mis atribuciones, las medidas inmediatas necesarias para limitar sus consecuencias y colaboraré con la revisión correspondiente.

17. No ocultamiento de incidentes

Reconozco que la detección y comunicación temprana de errores forma parte del funcionamiento normal del sistema de control.

No ocultaré deliberadamente un incidente, una exposición de información, una utilización no autorizada de herramientas o un error relevante producido durante el uso de inteligencia artificial.

La comunicación de buena fe no implicará por sí misma reconocimiento automático de responsabilidad.

18. Cumplimiento de procedimientos editoriales

No utilizaré inteligencia artificial para eludir verificaciones, revisiones editoriales, revisiones jurídicas, autorizaciones o cualquier otro control que resulte obligatorio para una actividad institucional.

La utilización de IA no modificará el nivel de aprobación requerido para una publicación.

Una automatización podrá ejecutar técnicamente una acción previamente autorizada, pero no sustituirá la decisión humana cuando esta sea exigida por el Manual.

19. Separación entre uso personal e institucional

Cuando utilice herramientas de inteligencia artificial para fines personales, procuraré mantener esa actividad separada de mis funciones institucionales.

No incorporaré información de Fundamedios a conversaciones personales simplemente porque la misma plataforma se encuentre disponible en ambos contextos.

La facilidad de copiar información entre cuentas no modifica las restricciones aplicables a su utilización.

20. Salida de la organización o cambio de funciones

Cuando termine mi participación en un proyecto, cambie de funciones o concluya mi relación con Fundamedios, colaboraré con la transferencia de los materiales institucionales que deban permanecer disponibles para la organización.

No conservaré en cuentas personales información confidencial, restringida o institucional cuya retención no se encuentre autorizada.

Cuando corresponda, transferiré la administración de proyectos, asistentes, automatizaciones, prompts, bases documentales u otros recursos institucionales que se encuentren bajo mi control.

21. Cooperación ante revisiones internas

Cuando se requiera investigar un incidente o revisar un proceso relacionado con inteligencia artificial, colaboraré razonablemente con la identificación de la información necesaria para reconstruir lo ocurrido.

Esta colaboración se limitará a las actividades y materiales institucionales pertinentes y deberá realizarse respetando las reglas aplicables de privacidad, confidencialidad y protección de derechos.

22. Reconocimiento de responsabilidad individual y de los límites de esta Declaración

Reconozco que el incumplimiento de las obligaciones asumidas mediante esta Declaración podrá ser evaluado conforme a la naturaleza, gravedad y consecuencias de la conducta y a las normas internas, contractuales o legales aplicables.

La existencia de un error generado por una herramienta de inteligencia artificial no determinará, por sí sola, mi responsabilidad.

La determinación correspondiente deberá considerar el procedimiento seguido, los controles aplicables, las instrucciones recibidas, las herramientas disponibles y mi actuación concreta.

De igual manera, la aceptación de esta Declaración no implica que Fundamedios quede exonerado de las obligaciones o responsabilidades que jurídicamente le correspondan.

23. Compromiso

Mediante la aceptación de esta Declaración manifiesto que comprendo que el uso institucional de inteligencia artificial implica responsabilidades relacionadas con la información, las fuentes, la seguridad, la verificación y la supervisión humana.

Me comprometo a utilizar estas herramientas de manera compatible con las finalidades de Fundamedios y con las reglas establecidas en el Manual.

Reconozco especialmente que:

la inteligencia artificial es una herramienta de asistencia y no sustituye mi obligación de aplicar los controles humanos, profesionales y de seguridad que correspondan a las actividades que se encuentren bajo mi responsabilidad.

CONSTANCIA DE ACEPTACIÓN

Nombre: _____

Cargo / función: _____

Proyecto / área: _____

Fecha: _____

Firma: _____

Anexo 2

Matriz de clasificación de información para uso con inteligencia artificial

La presente Matriz tiene por finalidad facilitar la aplicación práctica de las reglas de protección, confidencialidad, minimización y seguridad establecidas en el Manual.

Antes de introducir información en una herramienta de inteligencia artificial, la persona usuaria deberá identificar, en la medida razonablemente posible, **qué tipo de información está utilizando y qué nivel de protección requiere**.

La clasificación no dependerá únicamente del formato del documento ni de la circunstancia de que determinada información se encuentre disponible en Internet. Deberá atender principalmente a su contenido, contexto, finalidad y a las consecuencias que podrían producirse si fuera utilizada, almacenada, combinada o divulgada indebidamente.

Como regla general, se utilizarán cuatro niveles:

> **PÚBLICA → INTERNA → CONFIDENCIAL → RESTRINGIDA**

Estos niveles podrán complementarse con categorías especiales, como **datos personales**, **datos sensibles**, **fuentes protegidas** o **información jurídica reservada**, cuando la naturaleza del contenido así lo requiera.

---

1. Información pública

Se considerará **información pública** aquella destinada legítimamente al conocimiento general o que se encuentre disponible en fuentes abiertas sin restricciones relevantes para su utilización institucional.

Podrán encontrarse en esta categoría, entre otros ejemplos:

- publicaciones ya difundidas por Fundamedios;
- notas periodísticas publicadas;
- comunicados oficiales;
- información contenida en páginas web públicas;
- legislación publicada;
- sentencias de acceso público;
- bases de datos públicas;
- publicaciones abiertas en redes sociales;
- estadísticas oficiales;
- informes institucionales publicados;
- documentos académicos de acceso público; y
- cualquier otra información cuya utilización ordinaria no implique una expectativa razonable de confidencialidad.

Uso con IA

La información pública podrá utilizarse en herramientas institucionales autorizadas para búsquedas, análisis, clasificación, resumen, comparación, redacción y demás actividades permitidas por el Manual.

Sin embargo, la condición de pública **no elimina las obligaciones de verificación, finalidad, proporcionalidad y protección de datos**.

La recopilación masiva, combinación o perfilamiento de información pública sobre personas podrá generar riesgos superiores a los existentes en la fuente original y deberá evaluarse conforme al Manual.

Control requerido

Nivel ordinario de protección.

Deberá verificarse especialmente la exactitud, actualidad y procedencia de la información antes de utilizarla como sustento factual.

2. Información interna

Se considerará **información interna** aquella utilizada para el funcionamiento cotidiano de Fundamedios que no se encuentre destinada al público, pero cuya exposición accidental normalmente no produciría un perjuicio grave para personas o para la organización.

Podrán encontrarse en esta categoría:

- borradores ordinarios;
- agendas de trabajo no sensibles;
- documentos administrativos internos;
- formatos institucionales;
- metodologías no reservadas;
- notas de reuniones;
- cronogramas;
- materiales de capacitación;
- documentos de planificación general;
- instrucciones internas de trabajo;
- versiones preliminares de contenidos que no incluyan información sensible; y
- información organizativa cuya circulación se limite normalmente al personal o colaboradores.

Uso con IA

Podrá utilizarse exclusivamente en **cuentas y entornos institucionales autorizados**.

No deberá trasladarse sistemáticamente a cuentas personales.

Cuando el documento contenga elementos personales, confidenciales o de mayor sensibilidad, deberá aplicarse la clasificación superior correspondiente a esos elementos.

Control requerido

Acceso institucional + conservación dentro de repositorios autorizados.

La información interna podrá ser compartida entre quienes necesiten conocerla para cumplir sus funciones, pero no deberá entenderse como información de libre divulgación.

3. Información confidencial

Se considerará **información confidencial** aquella cuya divulgación, procesamiento externo o acceso no autorizado pueda producir un perjuicio relevante para Fundamedios, sus actividades, terceros o una investigación.

Podrán encontrarse en esta categoría:

a) investigaciones no publicadas;

b) documentos internos sensibles;

c) información recibida bajo acuerdos o expectativas razonables de confidencialidad;

d) comunicaciones internas relativas a decisiones editoriales sensibles;

e) estrategias institucionales no públicas;

f) borradores de investigaciones que contengan acusaciones o información todavía sujeta a contraste;

g) información jurídica interna;

h) contratos u otros documentos cuyo acceso se encuentre limitado;

i) bases de datos que contengan información no pública;

j) análisis internos de riesgos; y

k) cualquier material cuya exposición anticipada pueda comprometer el trabajo de Fundamedios.

Uso con IA

Solo podrá utilizarse en herramientas y entornos expresamente autorizados para procesar información confidencial.

Antes de introducirla deberá evaluarse si el mismo objetivo puede alcanzarse mediante:

- una versión depurada;
- eliminación de nombres;
- anonimización o seudonimización;
- selección de fragmentos;
- eliminación de anexos innecesarios; o
- utilización de información pública equivalente.

No deberá cargarse en cuentas personales ni en herramientas gratuitas o no evaluadas cuando estas no ofrezcan garantías suficientes.

Control requerido

Acceso limitado + entorno institucional autorizado + minimización + trazabilidad cuando corresponda.

Cuando la información se encuentre vinculada con un proyecto determinado, deberá procurarse que permanezca dentro de ese espacio y no dentro de memorias generales de la organización.

4. Información restringida

Se considerará **información restringida** aquella cuya exposición, pérdida, utilización indebida o acceso no autorizado pueda producir consecuencias graves o difícilmente reversibles.

Esta será la categoría de mayor protección.

Entre otros casos, podrá comprender:

- identidad de fuentes periodísticas confidenciales;
- información que permita razonablemente identificar una fuente protegida;
- datos cuya divulgación pueda poner en riesgo la seguridad física de una persona;
- información de víctimas, denunciantes o personas amenazadas cuando exista riesgo especial;
- credenciales, contraseñas, claves privadas, tokens y códigos de recuperación;
- información de seguridad de sistemas;
- documentación sometida a reserva legal;
- información jurídica especialmente protegida;
- datos personales sensibles cuando su tratamiento pueda generar un alto riesgo;
- bases de datos especialmente delicadas;
- información cuya exposición pueda comprometer una investigación de forma grave; y
- cualquier otra categoría que Fundamedios clasifique expresamente como restringida.

Uso con IA

La regla general será:

La información restringida no deberá introducirse en herramientas de inteligencia artificial de uso general.

Su procesamiento mediante IA solo podrá realizarse cuando:

1. exista una necesidad concreta y justificada;

2. la herramienta haya sido expresamente autorizada para ese tipo de información;

3. no exista una alternativa razonable menos riesgosa;

4. se hayan aplicado medidas de minimización y separación de información; y

5. exista el nivel de autorización correspondiente.

Cuando pueda realizarse la tarea sin revelar la identidad o elemento protegido, deberá utilizarse la versión depurada.

Control requerido

Acceso estrictamente limitado + autorización específica + protección técnica reforzada + trazabilidad + revisión del riesgo.

Determinada información restringida podrá quedar completamente fuera de los sistemas generativos aunque exista un entorno institucional.

5. Categorías especiales de información

La clasificación anterior deberá complementarse con algunas categorías que pueden aparecer dentro de cualquiera de los niveles y que requieren atención particular.

5.1. Datos personales

Se considerará dato personal cualquier información relacionada con una persona identificada o identificable conforme a la normativa aplicable.

Podrá incluir, según las circunstancias:

nombre — fotografía — voz — identificación — dirección — teléfono — correo electrónico — ubicación — información laboral — publicaciones — identificadores digitales — antecedentes — relaciones — opiniones atribuibles — información económica — y cualquier combinación de datos que permita identificar razonablemente a una persona.

Antes de utilizar datos personales mediante IA deberá determinarse:

¿Necesitamos este dato? ¿Necesitamos identificar a esta persona? ¿Podemos realizar la tarea con menos información? ¿La herramienta está autorizada para procesarlo?

La existencia de datos personales no convierte automáticamente toda información en restringida. El nivel dependerá de su naturaleza, finalidad, volumen, contexto y riesgo.

5.2. Datos sensibles o de protección reforzada

Cuando se utilicen categorías de datos sometidas a protección especial por la normativa aplicable, deberán aplicarse controles reforzados.

El tratamiento deberá limitarse a aquello que resulte realmente necesario para la finalidad legítima correspondiente.

Cuando sea posible, deberá evitarse proporcionar estos datos a la herramienta y trabajar con una versión anonimizada o seudonimizada.

Cuando la IA **infiera** una característica sensible de una persona, esa inferencia no deberá considerarse un hecho comprobado por el solo hecho de haber sido generada mediante análisis automatizado.

5.3. Fuentes confidenciales

Toda información capaz de identificar directa o indirectamente una fuente protegida deberá considerarse, como regla general, **restringida**.

Deberá mantenerse separada la identidad de la fuente del contenido que pueda ser legítimamente procesado.

Ejemplo:

En lugar de introducir:

“María Pérez, directora financiera de la institución X, me entregó este documento confidencial…”

podrá utilizarse, cuando resulte suficiente:

“Fuente A entregó el documento analizado.”

La información necesaria para relacionar internamente a “Fuente A” con la identidad real deberá mantenerse fuera del sistema utilizado cuando este no se encuentre autorizado para conocerla.

5.4. Información jurídica y procesal

Los documentos jurídicos deberán clasificarse atendiendo a su contenido.

Una sentencia publicada puede ser información pública.

Un expediente no público puede ser confidencial.

Una estrategia jurídica interna podrá ser restringida.

Una consulta legal preparada para Fundamedios podrá contener información de distintos niveles.

No deberá asumirse que todo documento jurídico tiene el mismo régimen únicamente porque se encuentra relacionado con un proceso legal.

5.5. Investigaciones periodísticas

Las investigaciones en curso deberán clasificarse atendiendo a su etapa y contenido.

Una investigación construida completamente sobre fuentes públicas puede comenzar como **interna**, pero convertirse en **confidencial** cuando aparecen hipótesis no publicadas, fuentes protegidas, denuncias o información cuya difusión anticipada pueda comprometer el trabajo.

La clasificación puede cambiar durante el desarrollo del proyecto.

El investigador deberá actualizar el nivel de protección cuando aparezcan nuevos elementos.

---

6. Regla de clasificación por el contenido más sensible

Cuando un archivo contenga información correspondiente a distintos niveles, deberá protegerse conforme al nivel más alto cuando no resulte razonablemente posible separar sus componentes.

Por ejemplo:

Un informe público al que se haya agregado una hoja con nombres de fuentes confidenciales no deberá tratarse en conjunto como información pública.

Una base de datos con cientos de registros abiertos puede requerir una clasificación superior si contiene además datos sensibles no públicos.

Una conversación de IA puede comenzar con información pública y posteriormente incorporar datos confidenciales. Desde ese momento deberán aplicarse los controles correspondientes a la nueva información.

Cuando sea posible, deberá separarse el componente sensible para evitar elevar innecesariamente la clasificación de todo el material.

---

7. Reclasificación

La clasificación no será permanente cuando cambien las circunstancias.

Una información podrá pasar de confidencial a pública después de su publicación.

Una información inicialmente interna podrá convertirse en restringida si posteriormente se determina que identifica a una fuente.

Una base aparentemente pública podrá requerir protección superior cuando su combinación permita generar perfiles sobre personas.

También podrá reducirse el nivel de protección cuando la información haya sido adecuadamente anonimizada, siempre que la anonimización impida razonablemente identificar nuevamente a las personas.

Toda reclasificación deberá atender al contenido real y no únicamente a una decisión administrativa anterior.

8. Matriz general de decisión

Clasificación	Ejemplos	Uso con IA	Cuenta personal	Acceso	Control principal
Pública	Normas, noticias, publicaciones oficiales, información publicada	Permitido en herramientas autorizadas	Evitar para trabajo institucional cuando exista entorno institucional	Ordinario	Verificación
Interna	Borradores ordinarios, metodologías, planificación	Entorno institucional	No recomendado / no cuando deba permanecer institucionalmente	Personal autorizado	Control institucional
Confidencial	Investigación no publicada, documentos internos sensibles	Solo herramientas aprobadas para ese nivel	No	Necesidad de conocer	Minimización + acceso restringido
Restringida	Fuentes protegidas, credenciales, datos de alto riesgo	Excepcional o prohibido según el caso	Prohibido	Estrictamente limitado	Autorización específica + controles reforzados

Esta matriz constituye una referencia general. Si existe una norma específica más protectora para determinada información, deberá aplicarse el estándar superior.

9. Preguntas antes de proporcionar información a una IA

Cuando exista duda sobre si determinada información puede incorporarse a una herramienta, deberá realizarse una evaluación sencilla.

PRIMERA PREGUNTA

¿La herramienta necesita realmente esta información para hacer la tarea?

Si la respuesta es no, no deberá proporcionarse.

SEGUNDA PREGUNTA

¿Existe información personal, confidencial, restringida o capaz de revelar una fuente?

Si la respuesta es sí, deberá revisarse la clasificación y las condiciones de tratamiento.

TERCERA PREGUNTA

¿Puedo eliminar o sustituir parte de la información sin afectar el objetivo?

Si es posible, deberá utilizarse la versión minimizada.

CUARTA PREGUNTA

¿La herramienta está autorizada para procesar este nivel de información?

Si no puede determinarse, el material no deberá cargarse hasta realizar la consulta correspondiente.

QUINTA PREGUNTA

¿Necesito que esta información permanezca disponible para la IA después de terminar la tarea?

Si la respuesta es no, deberá evitarse incorporarla a memorias o bases de conocimiento permanentes y aplicarse el procedimiento de conservación o eliminación correspondiente.

10. Ejemplos prácticos de clasificación

Caso A — Sentencia descargada del portal judicial

Clasificación probable: Pública.

Podrá utilizarse mediante IA institucional para resumen, búsqueda de argumentos o comparación, previa revisión del documento original antes de publicar conclusiones jurídicas.

Si el expediente adjunto contiene información adicional no pública o datos sensibles, deberá reclasificarse el material correspondiente.

Caso B — Entrevista con una fuente bajo reserva

Clasificación: Restringida.

La identidad deberá mantenerse fuera de herramientas generales. Si resulta necesario transcribir o analizar la entrevista, deberá utilizarse una herramienta específicamente autorizada o preparar previamente una versión que elimine elementos identificadores.

Caso C — Borrador de una nota todavía no publicada

Clasificación probable: Interna.

Si contiene acusaciones no publicadas, fuentes protegidas o información sensible, deberá elevarse a Confidencial o Restringida.

Caso D — Base pública de contratación estatal

Clasificación inicial: Pública.

Si Fundamedios agrega información no pública sobre personas, realiza perfiles o combina múltiples fuentes para obtener nuevas inferencias, deberá reevaluarse el nivel de riesgo y protección.

Caso E — Contraseña de acceso a una plataforma

Clasificación: Restringida.

No deberá introducirse en una herramienta de IA.

Caso F — Metodología de Ecuador Chequea

Clasificación: Interna o pública, dependiendo de que la versión correspondiente ya haya sido difundida.

Las versiones internas de trabajo deberán permanecer dentro del entorno institucional.

Caso G — Lista de periodistas amenazados con información de ubicación y contacto

Clasificación: Restringida.

Su procesamiento mediante IA deberá someterse a controles específicos y evitarse cuando no exista una necesidad estrictamente justificada.

Caso H — Nota periodística ya publicada por Fundamedios

Clasificación: Pública.

Podrá formar parte de una base de conocimiento institucional cuando su incorporación sea útil, sin perjuicio de mantener controles para evitar que información histórica sea presentada posteriormente como información actual.

11. Información destinada a la memoria institucional

La decisión de utilizar una información en una conversación de IA y la decisión de incorporarla permanentemente a una base de conocimiento institucional serán diferentes.

Para la memoria de uso general deberán preferirse contenidos:

- estables;
- validados;
- institucionalmente útiles;
- correctamente versionados;
- con bajo riesgo de generar exposición indebida; y
- cuya reutilización futura resulte previsible.

No deberán incorporarse automáticamente a la memoria general:

fuentes confidenciales; investigaciones restringidas; credenciales; datos sensibles innecesarios; información descartada; borradores no validados; información temporal; o documentos cuya reutilización futura pueda infringir obligaciones de confidencialidad.

Un documento podrá conservarse en el archivo institucional sin formar parte de la memoria activa de IA.

12. Responsabilidad de clasificación

La persona que incorpora información a una herramienta deberá realizar una evaluación inicial razonable de su naturaleza.

No se exigirá que cada usuario realice análisis jurídicos complejos antes de procesar cualquier documento.

Cuando la clasificación resulte evidente, deberá aplicarse directamente.

Cuando exista duda entre dos niveles y el contenido pueda producir consecuencias relevantes, deberá aplicarse provisionalmente **el nivel superior** y realizarse la consulta correspondiente.

Los responsables institucionales podrán emitir criterios adicionales para categorías de información utilizadas recurrentemente.

13. Regla final de clasificación

La clasificación deberá realizarse antes de que la información abandone su entorno original y sea incorporada a una herramienta de inteligencia artificial.

Una vez cargado un documento en una plataforma externa, determinadas consecuencias pueden ser difíciles o imposibles de revertir.

Por ello, la pregunta correcta deberá formularse **antes de cargar**, no después.

La regla final de esta Matriz será:

A mayor sensibilidad de la información, menor deberá ser su circulación, mayor deberá ser el control sobre la herramienta que la procesa y más fuerte deberá ser la justificación para utilizar inteligencia artificial con ella.

Anexo 3

Matriz institucional de riesgos para el uso de inteligencia artificial

La presente Matriz tiene por finalidad permitir que Fundamedios determine, antes y durante la utilización de inteligencia artificial, **qué nivel de riesgo presenta una actividad y qué controles deben aplicarse para realizarla**.

La evaluación no deberá concentrarse exclusivamente en la herramienta utilizada. Una misma plataforma puede representar un riesgo reducido cuando se emplea para corregir el estilo de un texto público y un riesgo elevado cuando se utiliza para analizar una investigación no publicada.

Por ello, el riesgo deberá determinarse considerando conjuntamente:

> **INFORMACIÓN + FINALIDAD + HERRAMIENTA + CAPACIDAD DE ACCIÓN + PERSONAS AFECTADAS + CONSECUENCIAS DEL ERROR**

La Matriz utilizará cuatro niveles generales:

> **BAJO → MEDIO → ALTO → CRÍTICO**

La clasificación deberá servir para determinar la intensidad de los controles y no para impedir innecesariamente la utilización de inteligencia artificial.

---

1. Criterios para evaluar el riesgo

Antes de clasificar una actividad deberán considerarse, cuando resulten pertinentes, los siguientes factores.

A. Sensibilidad de la información

Deberá determinarse si la actividad utiliza información:

- pública;
- interna;
- confidencial;
- restringida;
- personal;
- sensible;
- relacionada con fuentes protegidas; o
- capaz de producir riesgos para terceros.

La clasificación establecida en el **Anexo 2** constituirá el punto de partida.

B. Consecuencias de un error

Deberá analizarse qué ocurriría si la inteligencia artificial:

inventa información; interpreta incorrectamente un documento; confunde una persona; altera una cifra; atribuye falsamente una declaración; omite un elemento relevante; expone información; o ejecuta una acción equivocada.

Cuanto mayor sea el daño potencial, mayor deberá ser el nivel de control.

C. Destino del resultado

No presenta el mismo riesgo una respuesta utilizada únicamente como apoyo interno que una afirmación destinada a publicación.

Deberá distinguirse entre:

exploración interna → análisis → borrador → producto revisado → publicación → acción automatizada.

El acercamiento del resultado hacia una consecuencia externa normalmente incrementará el nivel de riesgo.

D. Nivel de autonomía

Deberá determinarse cuánto puede hacer la herramienta sin intervención humana.

Una IA que únicamente responde preguntas presenta un tipo de riesgo diferente de un agente que puede:

- consultar bases institucionales;
- enviar correos;
- modificar documentos;
- publicar contenidos;
- acceder a redes sociales;
- ejecutar código; o
- realizar acciones en otras plataformas.

A mayor autonomía, mayor deberá ser la supervisión.

E. Reversibilidad

Deberá considerarse si un error puede corregirse fácilmente.

Corregir internamente un resumen antes de utilizarlo es altamente reversible.

Publicar una acusación falsa, revelar la identidad de una fuente o enviar información confidencial a terceros puede producir consecuencias que no desaparecen simplemente eliminando posteriormente el contenido.

La baja reversibilidad incrementará el riesgo.

F. Escala

Una actividad automatizada puede multiplicar las consecuencias de un error.

No es equivalente utilizar IA para analizar un documento que utilizarla para procesar miles de registros, generar cientos de perfiles o producir contenidos automáticamente a gran escala.

La evaluación deberá considerar tanto la gravedad individual como la cantidad de personas, documentos o productos potencialmente afectados.

2. Riesgo bajo

Se considerará de **RIESGO BAJO** una actividad en la que la IA trabaje principalmente con información pública o institucional no sensible, el resultado permanezca bajo control humano y un error pueda detectarse y corregirse sin consecuencias relevantes.

Ejemplos

Podrán considerarse actividades de riesgo bajo:

- **corrección ortográfica de un texto;**
- **mejora de redacción;**
- **propuesta de títulos para un contenido ya validado;**
- **traducción preliminar de información pública;**
- **organización de una agenda;**
- **clasificación temática de publicaciones públicas;**
- **resumen de un documento público para comprensión interna;**
- **generación de ideas preliminares;** o
- **adaptación de formato de información previamente verificada.**

Controles

Las actividades de riesgo bajo podrán realizarse mediante el procedimiento ordinario.

La persona usuaria deberá revisar razonablemente el resultado antes de utilizarlo.

No será necesario documentar individualmente cada interacción, salvo que posteriormente forme parte de una actividad sometida a mayor nivel de control.

Regla

La IA puede asistir ampliamente; la persona mantiene control ordinario sobre el resultado.

3. Riesgo medio

Se considerará de **RIESGO MEDIO** una actividad en la que la IA intervenga en análisis sustantivos, trabaje con información interna o participe en la preparación de contenidos que posteriormente puedan llegar al público, pero donde los riesgos puedan controlarse mediante verificación y revisión humana ordinaria.

Podrán encontrarse en este nivel:

- elaboración de un primer borrador de una nota a partir de fuentes verificadas;
- resumen de varios documentos para preparar una investigación;
- comparación de declaraciones públicas;
- identificación preliminar de contradicciones;
- extracción de cifras de documentos;
- clasificación de información dentro de una investigación;
- elaboración de cronologías;
- transcripción de entrevistas no sensibles;
- análisis de tendencias sobre información pública;
- preparación de preguntas para una entrevista;

- adaptación de una publicación para diferentes formatos; y
- elaboración de contenidos internos a partir de documentación institucional ordinaria.

Controles requeridos

En estas actividades deberá existir:

1. Fuente identificable. La información factual deberá poder relacionarse con las fuentes correspondientes.

2. Revisión humana. El resultado deberá ser leído y evaluado antes de incorporarse al siguiente paso del proceso.

3. Verificación de elementos sustantivos. Nombres, cifras, fechas, citas y demás datos relevantes deberán comprobarse cuando vayan a utilizarse como hechos.

4. Entorno institucional. Cuando exista información interna, deberá utilizarse la cuenta o plataforma institucional correspondiente.

5. Control editorial posterior. Si el resultado llega a un producto público, deberá someterse a las revisiones establecidas en el Manual.

No será necesario conservar cada interacción exploratoria, pero sí aquellos elementos que resulten relevantes para reconstruir el producto cuando las reglas de trazabilidad así lo exijan.

Regla

La IA puede participar en el análisis y la producción, pero sus resultados sustantivos deben verificarse antes de avanzar.

4. Riesgo alto

Se considerará de **RIESGO ALTO** una actividad en la que un error, exposición o utilización incorrecta pueda producir consecuencias relevantes para Fundamedios, una investigación, una persona o el público.

La presencia de información confidencial, acusaciones, decisiones editoriales sensibles, datos personales significativos o automatizaciones con capacidad de actuar podrá elevar una actividad a este nivel.

Ejemplos

Podrán encontrarse en esta categoría:

- **análisis de investigaciones no publicadas;**
- **procesamiento de entrevistas con información confidencial;**
- **análisis de grandes conjuntos de datos personales;**
- **identificación de posibles relaciones entre personas o entidades;**
- **análisis jurídico utilizado para tomar decisiones institucionales relevantes;**
- **preparación de contenidos que contengan acusaciones o señalamientos;**
- **utilización de IA para evaluar material potencialmente falso o manipulado;**
- **conexión de IA con repositorios internos sensibles;**
- **automatizaciones capaces de modificar documentos institucionales;**
- **uso de sistemas capaces de enviar comunicaciones externas;** o
- **generación de contenidos audiovisuales que puedan confundirse con material auténtico.**

4.1. Controles reforzados

Las actividades de riesgo alto deberán incorporar, según corresponda:

Autorización de la herramienta. La plataforma deberá encontrarse aprobada para el tipo de información utilizada.

Acceso limitado. Solo deberán acceder quienes necesiten participar en la actividad.

Minimización. Deberá reducirse la información proporcionada al sistema a aquello necesario para cumplir la finalidad.

Trazabilidad. Deberán conservarse los elementos relevantes para reconstruir el proceso.

Verificación reforzada. Las afirmaciones importantes deberán contrastarse directamente con las fuentes originales.

Revisión humana especializada. Cuando el contenido requiera conocimientos periodísticos, jurídicos, técnicos o de otra naturaleza, deberá intervenir una persona con capacidad suficiente para evaluarlo.

Control antes de acciones externas. Ninguna automatización deberá ejecutar acciones relevantes sin la aprobación correspondiente.

Separación de espacios. Cuando sea posible, el proyecto deberá mantenerse separado de las memorias o repositorios generales.

Regla

La IA puede utilizarse, pero el procedimiento debe impedir que sus errores o accesos produzcan consecuencias externas sin controles reforzados.

---

5. Riesgo crítico

Se considerará de **RIESGO CRÍTICO** una actividad en la que una utilización incorrecta de inteligencia artificial pueda generar consecuencias graves, particularmente cuando puedan resultar comprometidas la seguridad de personas, la identidad de fuentes protegidas, información altamente sensible o derechos cuya afectación resulte difícilmente reversible.

La clasificación como crítica no significará necesariamente que toda utilización de IA se encuentre prohibida, pero deberá existir una **presunción de no utilización** hasta demostrar que el procesamiento es necesario y puede realizarse bajo condiciones suficientemente seguras.

Podrán encontrarse en este nivel:

- **identidad de fuentes confidenciales cuya exposición pueda generar represalias;**
- **información sobre ubicación o mecanismos de protección de personas amenazadas;**
- **credenciales y claves de seguridad;**
- **información cuya divulgación pueda producir un riesgo físico para una persona;**
- **datos sensibles a gran escala cuando exista riesgo significativo para los titulares;**
- **información sujeta a reserva cuya divulgación pueda generar consecuencias jurídicas graves;**
- **sistemas de IA con capacidad autónoma de publicar contenido institucional de alto impacto sin aprobación previa;**
- **automatizaciones con capacidad amplia para acceder, modificar y transmitir información restringida;** o
- **cualquier actividad cuya falla pueda producir un daño grave y difícilmente reversible.**

6. Controles para actividades de riesgo crítico

Antes de utilizar IA en una actividad clasificada como crítica deberá determinarse expresamente:

1. Necesidad

¿Existe una razón suficiente para utilizar inteligencia artificial?

La simple conveniencia o ahorro de tiempo no deberá considerarse, por sí solo, justificación suficiente cuando puedan utilizarse mecanismos menos riesgosos.

2. Alternativa

¿Puede realizarse la tarea sin proporcionar la información crítica?

Deberá considerarse anonimización, seudonimización, fragmentación, eliminación de datos, procesamiento local u otros mecanismos.

3. Herramienta

¿La plataforma ha sido expresamente evaluada para procesar ese tipo de información?

Una autorización general para utilizar una herramienta no implicará autorización para actividades críticas.

4. Acceso

¿Quién podrá acceder al proyecto, conversación, repositorio o resultado?

El acceso deberá limitarse estrictamente.

5. Persistencia

¿La información quedará almacenada, incorporada a memoria o disponible para usos posteriores?

Deberá evitarse toda persistencia innecesaria.

6. Consecuencia

¿Qué ocurriría si el sistema falla o la información es expuesta?

La respuesta deberá determinar si el riesgo residual resulta aceptable.

7. Autorización

La utilización deberá contar con autorización institucional previa de la persona o función competente.

Cuando el riesgo residual continúe siendo excesivo, **la actividad no deberá realizarse mediante IA**.

7. Factores que elevan automáticamente el nivel de revisión

Determinadas circunstancias deberán provocar una revisión adicional aunque inicialmente la actividad parezca de menor riesgo.

Entre ellas:

- participación de menores de edad;
- posibles víctimas de violencia;
- personas bajo protección;
- fuentes confidenciales;
- grandes volúmenes de datos personales;
- información sobre salud u otras categorías sensibles;
- investigaciones que puedan producir acusaciones graves;
- material cuya autenticidad sea controvertida;
- utilización de reconocimiento biométrico;
- generación de imágenes, voz o video realistas de personas;
- decisiones automatizadas que puedan afectar significativamente a terceros;
- conexiones con sistemas institucionales;
- ejecución automática de acciones;
- publicación automática;
- cambios importantes en el funcionamiento de la herramienta; y
- cualquier situación en la que la persona usuaria no comprenda razonablemente qué hará el sistema con la información.

La presencia de uno de estos factores no determinará necesariamente que el uso sea crítico, pero impedirá tratarlo automáticamente como una actividad ordinaria.

8. Factores de riesgo editorial

Debido a la naturaleza de las actividades de Fundamedios, deberá realizarse una evaluación específica cuando la IA intervenga en contenidos destinados al público.

El riesgo editorial aumentará cuando el producto incluya:

- **acusaciones;**
- **atribución de conductas ilegales;**
- **información que pueda afectar reputación;**
- **citas textuales;**
- **resultados de investigaciones;**
- **información sobre personas vulnerables;**
- **interpretaciones jurídicas;**
- **cifras determinantes para la conclusión;**
- **material cuya autenticidad se encuentre bajo análisis;** o
- **afirmaciones que dependan de una única fuente.**

En estos casos, la fluidez o aparente seguridad con la que la IA redacte una afirmación no deberá influir en el nivel de evidencia exigido para publicarla.

9. Riesgo de alucinación y distorsión

Toda utilización de IA generativa deberá considerar la posibilidad de que el sistema produzca información incorrecta con apariencia de certeza.

Este riesgo deberá considerarse especialmente alto cuando se soliciten:

- referencias jurídicas;
- jurisprudencia;
- citas textuales;
- estadísticas;
- nombres de documentos;
- fechas históricas específicas;
- antecedentes de personas;
- referencias bibliográficas;
- enlaces;
- detalles que no se encuentren en el material proporcionado; o
- reconstrucciones de acontecimientos con información incompleta.

Cuando una respuesta contenga uno de estos elementos, la herramienta deberá entenderse como **mecanismo de localización o asistencia**, no como prueba de que el elemento existe.

10. Riesgo de automatización

Una actividad deberá recibir mayor nivel de control cuando la IA pueda ejecutar acciones sin que una persona revise individualmente cada resultado.

Deberán diferenciarse tres niveles:

IA asistencial

La herramienta propone y una persona decide.

Ejemplo: redacta un borrador.

IA semiautomatizada

La herramienta realiza parte del proceso automáticamente, pero existe una instancia humana antes de producir consecuencias externas.

Ejemplo: genera y prepara una publicación, pero una persona debe aprobarla.

IA automatizada

La herramienta puede producir directamente una consecuencia externa.

Ejemplo: publica, envía, modifica o elimina información sin aprobación inmediata.

Fundamedios deberá preferir los modelos **asistenciales o semiautomatizados** para actividades editoriales y de alto impacto.

La automatización completa deberá reservarse para actividades suficientemente previsibles y de riesgo controlado.

11. Riesgo acumulativo

Una serie de actividades individualmente moderadas puede producir un riesgo elevado cuando se combinan.

Por ejemplo:

recopilar información pública sobre una persona → combinar múltiples bases → inferir relaciones → generar un perfil → producir una evaluación automática.

Cada etapa puede parecer limitada de manera aislada, pero el resultado conjunto puede contener información que no existía explícitamente en ninguna de las fuentes originales.

La evaluación deberá considerar **el proceso completo y no únicamente cada operación individual**.

12. Cambio del nivel de riesgo durante el trabajo

La clasificación inicial no será definitiva.

Una investigación puede comenzar utilizando únicamente información pública y posteriormente recibir documentación confidencial.

Un borrador puede convertirse en una publicación de alto impacto.

Una herramienta inicialmente utilizada para consulta puede posteriormente conectarse a un repositorio.

Una automatización puede adquirir nuevas capacidades.

Cuando cambien las condiciones, deberá revisarse el nivel de riesgo y aplicarse los controles correspondientes antes de continuar.

13. Matriz resumida de riesgos

Nivel	Característic as	Ejemplo	Uso de IA	Control humano
BAJO	Información pública, efectos fácilmente reversibles	Corrección de estilo	Permitido	Revisión ordinaria
MEDIO	Análisis sustantivo o contenido preparatorio	Borrador basado en fuentes verificadas	Permitido	Verificación antes de uso
ALTO	Información confidencial o consecuencias relevantes	Investigación no publicada	Permitido con controles reforzados	Revisión especializada
CRÍTICO	Posibilidad de daño grave o irreversible	Identidad de fuente en riesgo	Excepcional o prohibido	Autorización específica

14. Árbol práctico de decisión

Antes de iniciar una actividad, la persona usuaria podrá aplicar la siguiente secuencia:

1. ¿Estoy utilizando únicamente información pública o interna ordinaria?

Si la respuesta es **sí**, continuar con la evaluación.

Si es **no**, revisar si existe información confidencial o restringida.

2. ¿El resultado puede llegar al público o afectar a una persona?

Si es **sí**, aumentar el nivel de revisión.

3. ¿La IA puede ejecutar una acción fuera de la conversación?

Si es **sí**, aplicar controles de autorización.

4. ¿Un error puede producir consecuencias importantes?

Si es **sí**, clasificar al menos como riesgo medio o alto según la gravedad.

5. ¿Existe información cuya exposición pueda producir un daño grave?

Si es **sí**, evaluar como riesgo crítico.

6. ¿Puedo reducir el riesgo eliminando información o limitando las capacidades de la herramienta?

Si es **sí**, aplicar esas medidas antes de comenzar.

7. ¿Después de aplicar los controles el riesgo sigue siendo excesivo?

Si es **sí**, no utilizar IA para esa actividad.

---

15. Registro de evaluaciones de alto riesgo

No será necesario documentar formalmente cada utilización cotidiana de inteligencia artificial.

Las actividades clasificadas como **alto riesgo** que tengan carácter recurrente podrán contar con una evaluación general previamente aprobada.

Por ejemplo, si Fundamedios establece un procedimiento seguro para analizar determinados documentos confidenciales dentro de una plataforma institucional, no será necesario aprobar individualmente cada documento cuando las condiciones permanezcan iguales.

En cambio, las actividades críticas o aquellas que introduzcan una capacidad sustancialmente nueva deberán documentarse específicamente.

El registro podrá indicar:

actividad evaluada;

nivel de riesgo;

información utilizada;

herramienta;

controles establecidos;

riesgo residual;

persona o función que autorizó; y

fecha de revisión.

16. Riesgo residual

La aplicación de controles no elimina completamente el riesgo.

Después de establecer medidas deberá considerarse qué riesgo permanece.

Ejemplo:

Una investigación confidencial puede procesarse dentro de una cuenta institucional, con acceso restringido, sin nombres de fuentes y con historial controlado.

Estas medidas reducen significativamente el riesgo, pero no lo convierten necesariamente en cero.

Fundamedios deberá decidir si el **riesgo residual** resulta razonable frente a la utilidad de la actividad.

Cuando las consecuencias potenciales sean extraordinariamente graves, incluso una probabilidad reducida podrá justificar no utilizar IA.

17. Revisión de la Matriz

Los criterios de esta Matriz deberán revisarse conforme evolucione la utilización institucional de inteligencia artificial.

Los incidentes deberán servir para comprobar si determinadas actividades fueron clasificadas incorrectamente.

Si una categoría de uso produce reiteradamente errores o incidentes, podrá elevarse su nivel de control.

Si la experiencia y las mejoras tecnológicas permiten demostrar que determinados riesgos se encuentran adecuadamente mitigados, podrán ajustarse los procedimientos sin reducir las garantías esenciales.

18. Regla final de gestión del riesgo

La finalidad de esta Matriz no será decidir simplemente entre **“usar IA”** o **“no usar IA”**.

Su función será determinar:

qué puede hacerse, con qué información, mediante qué herramienta, bajo qué condiciones y con qué nivel de supervisión.

El nivel de control deberá aumentar a medida que aumenten la sensibilidad de la información, la autonomía de la herramienta y las consecuencias potenciales de un error.

La regla final será:

Fundamedios utilizará inteligencia artificial cuando su utilidad resulte compatible con un nivel de riesgo razonablemente controlado. Cuando los controles disponibles no permitan reducir un riesgo grave a un nivel aceptable, la actividad deberá realizarse sin inteligencia artificial o mediante un procedimiento alternativo más seguro.

Anexo 4

Lista de comprobación para investigación y levantamiento de información mediante inteligencia artificial

La presente lista de comprobación tiene por finalidad trasladar las disposiciones del Manual al trabajo cotidiano de investigación, monitoreo, búsqueda, recopilación, sistematización y análisis inicial de información realizado por Fundamedios, Ecuador Chequea, Liberties 360 y las demás iniciativas, proyectos y plataformas de la organización.

No constituye un formulario que deba completarse obligatoriamente para cada consulta realizada a una herramienta de inteligencia artificial. Su función es proporcionar una **secuencia práctica de control** para aquellas actividades en las que la IA intervenga de manera sustantiva en el levantamiento o análisis de información.

La intensidad de los controles deberá corresponder al riesgo de la actividad.

Una búsqueda exploratoria sobre información pública no requerirá el mismo nivel de documentación que una investigación que involucre fuentes confidenciales, datos personales o información que posteriormente pueda sustentar una acusación.

La lógica general será:

> **DEFINIR → CLASIFICAR → BUSCAR → DOCUMENTAR → VERIFICAR → ORGANIZAR → ANALIZAR → CONSERVAR**

1. Antes de comenzar: definir qué se está buscando

Antes de utilizar inteligencia artificial para investigar deberá existir una pregunta, objetivo o necesidad de información suficientemente identificable.

No deberá iniciarse una recopilación indiscriminada de información simplemente porque una herramienta permita hacerlo.

La persona responsable deberá poder responder, al menos de manera general:

¿Qué necesito conocer?

¿Para qué necesito conocerlo?

¿Qué información sería relevante para responder esa pregunta?

¿Qué fuentes podrían contenerla?

¿Qué nivel de certeza necesitaré antes de utilizarla?

Una investigación podrá evolucionar y descubrir elementos que inicialmente no habían sido previstos. La definición inicial de objetivos no deberá impedir la exploración legítima, pero sí proporcionar una referencia que evite búsquedas innecesariamente amplias o recopilaciones desproporcionadas de información.

2. Determinar el papel que tendrá la IA

Antes de iniciar una investigación deberá identificarse qué función se pretende asignar a la herramienta.

La IA podrá utilizarse, entre otras actividades, para:

- localizar posibles fuentes;
- formular estrategias de búsqueda;
- identificar documentos;
- resumir información;
- extraer datos;
- clasificar documentos;
- comparar versiones;
- construir cronologías;
- detectar inconsistencias;
- identificar preguntas pendientes;
- relacionar información previamente recopilada;
- organizar grandes volúmenes de material;
- traducir documentos;
- transcribir contenido;
- sugerir líneas adicionales de investigación; o
- asistir en el análisis preliminar.

La función deberá encontrarse suficientemente delimitada para determinar posteriormente qué resultados necesitan verificación.

La IA puede ayudar a encontrar una pista. La pista no se convierte en hecho hasta ser verificada.

3. Identificar el nivel de riesgo

Antes de proporcionar información institucional a la herramienta deberá revisarse la **Matriz Institucional de Riesgos**.

La evaluación inicial podrá realizarse mediante cuatro preguntas:

¿La investigación utiliza información pública?

Si utiliza únicamente información pública y el resultado permanecerá como exploración interna, normalmente podrá utilizarse el procedimiento ordinario.

¿Existe información interna o confidencial?

Si existe, deberá comprobarse que la herramienta se encuentre autorizada para procesarla.

¿Existen fuentes protegidas, datos sensibles o información capaz de producir un daño grave?

Si existe, deberán aplicarse los controles correspondientes a información restringida o riesgo crítico.

¿La investigación puede producir una publicación o decisión de alto impacto?

Si la respuesta es afirmativa, deberá aumentarse desde el inicio el nivel de trazabilidad y verificación.

4. Clasificar la información antes de cargarla

Antes de cargar documentos, bases de datos, imágenes, audios u otros materiales deberá determinarse su clasificación conforme al **Anexo 2**.

La persona deberá identificar si la información es:

PÚBLICA / INTERNA / CONFIDENCIAL / RESTRINGIDA

Cuando exista duda entre dos niveles y la diferencia pueda ser relevante para la seguridad, deberá utilizarse provisionalmente el nivel superior.

No deberá esperarse a que la información haya sido proporcionada a la herramienta para analizar si debía haber sido protegida.

5. Aplicar minimización

Una vez determinada la información necesaria, deberá revisarse si es posible reducirla antes de proporcionarla a la IA.

Podrá utilizarse, según corresponda:

- **eliminación de nombres;**
- **seudónimos;**
- **eliminación de números de identificación;**
- **supresión de datos de contacto;**
- **eliminación de metadatos;**
- **selección únicamente de las páginas necesarias;**
- **fragmentación del documento;**
- **separación entre identidad y contenido;** o
- **sustitución de información real por categorías cuando la identidad no sea necesaria.**

La pregunta práctica será:

¿Puede la IA realizar correctamente esta tarea con menos información?

Si la respuesta es sí, deberá proporcionarse menos información.

6. Utilizar la herramienta adecuada

Antes de comenzar deberá comprobarse que la herramienta utilizada se encuentre autorizada para la actividad y para el nivel de información correspondiente.

La selección no deberá realizarse únicamente por conveniencia.

Deberá considerarse:

capacidad de búsqueda;

acceso a fuentes;

tratamiento de archivos;

protección de información;

memoria;

integraciones;

trazabilidad;

capacidad para proporcionar enlaces o referencias; y

limitaciones conocidas.

Una herramienta excelente para redacción puede no ser la mejor herramienta para investigación documental.

Una herramienta capaz de buscar información puede no encontrarse autorizada para recibir documentos confidenciales.

7. Diseñar correctamente la instrucción de investigación

Los prompts utilizados para investigación deberán procurar delimitar claramente la tarea.

Cuando corresponda, la instrucción deberá especificar:

OBJETO

Qué se busca.

PERÍODO

Qué fechas deben cubrirse.

ÁMBITO

Qué personas, instituciones, lugares, materias o documentos se encuentran comprendidos.

FUENTES

Qué tipos de fuentes deben priorizarse.

EXCLUSIONES

Qué información no resulta relevante o qué fuentes no deberán utilizarse.

FORMATO

Cómo deberá organizarse el resultado.

EVIDENCIA

Qué referencias deberá proporcionar la herramienta para permitir la verificación.

INCERTIDUMBRE

Cómo deberá identificar aquello que no pudo comprobar.

Un prompt de investigación no deberá limitarse, cuando la precisión sea relevante, a instrucciones generales como:

“Busca todo sobre este tema.”

Deberá proporcionar parámetros suficientes para reducir resultados irrelevantes y permitir evaluar posteriormente la cobertura de la búsqueda.

8. Jerarquía de fuentes

La IA no deberá decidir por sí sola qué fuente merece confianza únicamente por aparecer primero en una búsqueda.

La investigación deberá priorizar, según la naturaleza del asunto:

fuentes primarias;

documentos oficiales;

bases institucionales;

registros públicos;

legislación;

sentencias;

informes originales;

documentos académicos;

medios reconocidos;

organizaciones especializadas; y

otras fuentes cuya procedencia pueda evaluarse.

Las fuentes secundarias podrán utilizarse para contextualizar, localizar documentos o identificar líneas de investigación.

Una publicación secundaria que cite un documento primario deberá conducir, cuando resulte posible, a la revisión del documento original.

9. La IA no será citada como fuente de los hechos

Las respuestas generadas por una herramienta de inteligencia artificial no deberán registrarse como fuente factual de una investigación.

No deberá indicarse, por ejemplo:

“Fuente: ChatGPT”.

La fuente deberá ser el documento, base de datos, declaración, publicación, entrevista, registro o evidencia que sustente la información.

La herramienta deberá funcionar como **intermediaria de búsqueda, procesamiento o análisis**, no como sustituto de la evidencia.

Cuando la IA no pueda identificar el origen de una afirmación, esta deberá considerarse no verificada.

10. Registrar la procedencia de la información

Durante el levantamiento deberá procurarse que cada hallazgo relevante conserve una relación con su fuente.

El registro podrá incluir:

título del documento;

autor o institución;

fecha;

enlace o ubicación;

página, sección o fragmento relevante;

fecha de consulta cuando corresponda;

observaciones; y

estado de verificación.

Cuando se trabaje con gran cantidad de información podrá utilizarse una tabla o repositorio estructurado.

Ejemplo:

Hallazgo	Fuente	Ubicación	Estado	Observación
Persona X ocupó el cargo Y	Resolución institucional	pág. 4	Verificado	Fuente primaria
Empresa A estaría vinculada con B	Nota periodística	enlace	Pendiente	Buscar registro societario
Cifra de USD X millones	Informe oficial	tabla 6	Verificado	Confirmar año

La estructura podrá adaptarse a cada investigación.

11. Diferenciar hechos, afirmaciones, hipótesis e inferencias

Durante el levantamiento deberá evitarse que información con distinto nivel de certeza quede mezclada dentro de un mismo conjunto de datos.

Podrá utilizarse una clasificación como:

HECHO VERIFICADO

Existe evidencia suficiente que permite sostener la afirmación.

AFIRMACIÓN ATRIBUIDA

Una persona, institución o fuente afirma algo, pero ello no significa necesariamente que el contenido haya sido comprobado.

HALLAZGO PENDIENTE

Existe información potencialmente relevante que requiere confirmación.

HIPÓTESIS DE INVESTIGACIÓN

Existe una posible explicación o relación que deberá investigarse.

INFERENCIA DE IA

La herramienta identificó una posible conclusión o patrón que todavía no ha sido verificado por una persona.

Esta separación será particularmente importante cuando posteriormente se utilice la información para redactar.

12. Verificar citas textuales

Las citas producidas por inteligencia artificial deberán recibir control especial.

Cuando una herramienta atribuya a una persona una frase entre comillas, la cita deberá verificarse contra:

- grabación;
- transcripción;
- documento original;
- publicación original; o
- otra fuente primaria adecuada.

Si no puede localizarse la cita, no deberá utilizarse como cita textual.

Una IA puede resumir correctamente una idea y, al mismo tiempo, fabricar la formulación exacta de una frase.

Por ello:

Las comillas exigen fuente original.

13. Verificar cifras y estadísticas

Toda cifra relevante deberá poder rastrearse hasta una fuente.

La revisión deberá comprobar, cuando corresponda:

- **cantidad;**
- **unidad;**
- **moneda;**
- **período;**
- **universo;**
- **metodología;**
- **fecha de actualización;** y
- **contexto.**

Un porcentaje correcto puede resultar engañoso si se aplica al período equivocado.

Una cifra correcta puede cambiar completamente de significado si corresponde a miles y no a millones.

La IA podrá ayudar a extraer o comparar cifras, pero la comprobación deberá realizarse contra el material original cuando el dato sea sustantivo.

14. Verificación de normas, sentencias y documentos oficiales

Cuando una investigación involucre información jurídica, la IA deberá utilizarse con especial cautela.

Toda referencia relevante deberá comprobarse directamente.

Deberán verificarse:

- **existencia de la norma o decisión;**
- **numeración;**

- **fecha;**
- **autoridad que la emitió;**
- **vigencia;**
- **texto aplicable;**
- **contexto;** y
- **alcance real de la disposición citada.**

No deberá incorporarse una norma o sentencia únicamente porque la herramienta proporcione un número, nombre o cita aparentemente plausible.

Cuando el contenido jurídico sea determinante para una publicación o actuación institucional, deberá aplicarse la revisión especializada correspondiente.

15. Control de enlaces

Los enlaces proporcionados por IA deberán comprobarse antes de incorporarse a una investigación.

Deberá verificarse que:

existan;

conduzcan al documento indicado;

correspondan a la institución o fuente señalada;

no sean versiones manipuladas; y

contengan efectivamente la información atribuida.

Cuando sea relevante preservar evidencia de una fuente que puede modificarse o desaparecer, deberá utilizarse el mecanismo institucional correspondiente para conservarla.

16. Información procedente de redes sociales

La IA podrá ayudar a localizar, organizar o analizar contenido de redes sociales.

Sin embargo, deberá distinguirse entre:

que una publicación exista y **que aquello que afirma la publicación sea verdadero.**

La existencia de una publicación podrá verificarse mediante la fuente original.

El contenido factual de la publicación podrá requerir evidencia adicional.

Cuando resulte relevante deberá comprobarse también:

autoría de la cuenta;

fecha;

contexto;

posibles ediciones;

contenido completo; y

autenticidad del material audiovisual asociado.

17. Información generada por otras inteligencias artificiales

El hecho de que un contenido haya sido generado, resumido o analizado previamente por otra herramienta de IA deberá considerarse al evaluar su confiabilidad.

Una respuesta generada por un sistema no deberá utilizarse para "confirmar" otra respuesta generada por un sistema diferente si ambas carecen de fuentes independientes.

Dos inteligencias artificiales que repiten la misma afirmación no equivalen a dos fuentes.

La corroboración deberá realizarse mediante evidencia independiente.

18. Búsqueda de información contradictoria

Una investigación asistida por IA no deberá limitarse a buscar información que confirme la hipótesis inicial.

Cuando la materia lo justifique, deberán realizarse búsquedas orientadas a encontrar:

versiones alternativas;

desmentidos;

explicaciones distintas;

documentos contradictorios;

errores en la hipótesis inicial;

información que favorezca a la persona investigada; y

evidencia que pueda modificar la conclusión.

Podrá utilizarse expresamente la IA para realizar una revisión adversarial:

"Identifique evidencia que contradiga esta hipótesis y señale qué elementos deberían verificarse antes de sostenerla."

El resultado seguirá requiriendo verificación.

19. Evitar sesgo de confirmación automatizado

La forma en que se formula una pregunta puede orientar la respuesta de la herramienta.

No será equivalente preguntar:

"¿Qué pruebas existen de que X cometió esta irregularidad?"

que preguntar:

"¿Qué información verificable existe sobre la posible participación de X en este hecho? Identifique evidencia favorable, desfavorable y elementos que no puedan determinarse."

Cuando todavía no exista una conclusión, los prompts deberán evitar formular como hecho aquello que únicamente constituye una hipótesis.

20. Construcción de cronologías

La IA podrá resultar especialmente útil para ordenar hechos cronológicamente.

Cuando se utilice para este propósito deberá solicitarse, en la medida posible:

fecha → acontecimiento → fuente → nivel de certeza.

Ejemplo:

Fecha	Hecho	Fuente	Estado
12/03/2026	Se emitió la resolución	Documento oficial	Verificado
18/03/2026	Medio reportó reunión	Medio X	Verificado como publicación
20/03/2026	Supuesta transferencia	IA detectó referencia	Pendiente

La cronología no deberá eliminar incertidumbres simplemente para producir una narrativa más ordenada.

21. Análisis de relaciones

La inteligencia artificial podrá utilizarse para identificar posibles relaciones entre personas, empresas, instituciones, acontecimientos o documentos.

Las relaciones deberán clasificarse conforme a su evidencia.

Deberá distinguirse entre:

relación documentada;

coincidencia;

correlación;

inferencia; y

mera posibilidad.

No deberá transformarse automáticamente una coincidencia de nombres, direcciones, fechas, accionistas, fotografías o eventos en una afirmación de relación causal o ilícita.

La IA podrá señalar patrones.

La investigación deberá demostrar qué significan.

22. Identificación de personas

Cuando exista riesgo de homonimia, la identidad deberá confirmarse mediante información adicional.

Un nombre idéntico no demuestra que dos registros correspondan a la misma persona.

Cuando sea relevante podrán utilizarse elementos como:

cargo; institución; ciudad; período; documentos oficiales; identificadores públicos legítimamente disponibles; y otros elementos de corroboración.

La IA no deberá resolver una ambigüedad de identidad mediante una inferencia presentada como certeza.

23. Traducciones

La IA podrá utilizarse para traducir información durante una investigación.

Cuando una palabra, expresión o formulación tenga importancia sustantiva, jurídica o probatoria, deberá revisarse el texto original.

Las traducciones automáticas podrán utilizarse para comprensión y trabajo preliminar, pero no deberán sustituir la comprobación especializada cuando una diferencia lingüística pueda modificar una conclusión.

24. Transcripciones

Las transcripciones automatizadas deberán considerarse borradores hasta ser revisadas cuando su contenido sea relevante.

Especial atención deberá prestarse a:

- nombres propios;
- cifras;
- negaciones;
- fechas;
- tecnicismos;
- palabras pronunciadas de forma poco clara;
- cambios de interlocutor; y
- fragmentos utilizados como citas.

Cuando una cita vaya a publicarse deberá comprobarse contra el audio o video original cuando este se encuentre disponible.

25. Imágenes y material audiovisual

La IA podrá utilizarse para analizar imágenes o material audiovisual, pero las conclusiones deberán ajustarse a aquello que realmente puede determinarse.

No deberá afirmarse la identidad de una persona, lugar, fecha, objeto o acontecimiento únicamente por una inferencia automática cuando ello requiera evidencia adicional.

Cuando la autenticidad sea relevante deberán aplicarse los mecanismos de verificación disponibles y revisar, cuando corresponda:

fuente original;

metadatos;

publicaciones anteriores;

contexto;

ediciones;

manipulaciones; y

posible generación sintética.

26. Control de actualidad

Las herramientas de IA pueden trabajar con información desactualizada.

Toda investigación sobre cargos, autoridades, legislación vigente, estadísticas recientes, situaciones en desarrollo u otros elementos sujetos a cambio deberá incluir verificación temporal.

La persona deberá preguntarse:

¿Esta información sigue siendo cierta hoy o solamente fue cierta en algún momento?

Cuando el período sea importante deberá incluirse expresamente en el prompt y en la verificación.

27. Organización de los resultados

Los resultados útiles deberán trasladarse progresivamente a los repositorios institucionales correspondientes.

No deberá mantenerse una investigación completa únicamente dentro de una conversación con una herramienta.

La organización podrá utilizar, según el proyecto:

carpetas;

tablas;

bases documentales;

cronologías;

fichas de fuentes;

matrices de hallazgos;

repositorios de evidencia; o

otros sistemas adecuados.

La estructura deberá permitir que otra persona autorizada pueda comprender razonablemente el origen y estado de la información.

28. Estado de verificación

Cuando se manejen numerosos hallazgos podrá utilizarse un sistema visual o textual sencillo:

VERIFICADO — existe evidencia suficiente.

PENDIENTE — requiere comprobación adicional.

DESCARTADO — la información resultó incorrecta o irrelevante.

CONTRADICTORIO — existen fuentes relevantes incompatibles.

NO DETERMINADO — no existe evidencia suficiente para concluir.

Los resultados de IA no deberán pasar automáticamente a la categoría “verificado”.

29. Información descartada

Cuando una hipótesis o dato haya sido descartado por resultar incorrecto, deberá evitarse que permanezca mezclado con información validada.

En investigaciones relevantes podrá conservarse registro de que determinada hipótesis fue revisada y descartada, especialmente cuando exista riesgo de que reaparezca posteriormente.

La información descartada no deberá incorporarse a una base de conocimiento general como si fuera verdadera.

30. Cierre de la fase de levantamiento

Antes de utilizar la información recopilada para producir una nota, alerta, informe, verificación u otro producto, deberá realizarse un control de cierre.

La persona responsable deberá poder determinar:

FUENTES

¿Sabemos de dónde provienen los principales hechos?

VERIFICACIÓN

¿Los elementos sustantivos han sido comprobados?

INCERTIDUMBRES

¿Está claramente identificado aquello que todavía no sabemos?

CONTRADICCIONES

¿Se revisaron las versiones relevantes que contradicen la hipótesis?

DATOS PERSONALES

¿La información personal recopilada sigue siendo necesaria?

FUENTES PROTEGIDAS

¿Las identidades y elementos identificadores se encuentran adecuadamente separados?

ARCHIVO

¿Los documentos relevantes se encuentran almacenados en el repositorio institucional correspondiente?

TRAZABILIDAD

¿Podemos reconstruir razonablemente cómo llegamos a los hallazgos principales?

Si alguna de estas preguntas no puede responderse satisfactoriamente, deberá evaluarse si la investigación se encuentra preparada para pasar a la fase de producción.

31. Lista rápida de comprobación

Antes de cerrar el levantamiento de información, podrá utilizarse la siguiente lista:

☐ Definí qué necesitaba investigar.

☐ Determiné qué función tendría la IA.

☐ Clasifiqué la información utilizada.

☐ Utilicé una herramienta autorizada.

☐ Eliminé información innecesaria cuando correspondía.

☐ Conservé las fuentes originales.

☐ Separé hechos de hipótesis e inferencias.

☐ Verifiqué nombres, fechas, cifras y citas relevantes.

☐ Comprobé normas y documentos oficiales cuando correspondía.

☐ Revisé los enlaces utilizados.

☐ Busqué información contradictoria cuando era pertinente.

☐ Identifiqué claramente los hallazgos pendientes.

☐ Protegí fuentes y datos sensibles.

☐ Organicé los materiales en el repositorio institucional.

☐ Puedo explicar de dónde provienen los hechos principales.

☐ La información se encuentra preparada para pasar a la fase de producción.

32. Regla final del levantamiento de información

La inteligencia artificial puede ampliar considerablemente la capacidad de Fundamedios para localizar, procesar y relacionar información. Esa capacidad no deberá confundirse con la producción automática de conocimiento confiable.

Una herramienta puede encontrar rápidamente cientos de referencias y, aun así, equivocarse en el hecho central de una investigación.

Por ello, el valor de la IA en esta fase dependerá de la calidad del proceso que la rodea: **una pregunta bien definida, fuentes identificables, clasificación adecuada, trazabilidad y verificación humana.**

La regla final de esta lista de comprobación será:

Ningún dato será considerado verdadero por el solo hecho de haber sido encontrado, resumido, relacionado o afirmado por una inteligencia artificial. La información adquirirá valor institucional cuando pueda vincularse con evidencia verificable y haya sido sometida al nivel de revisión que corresponda a su riesgo.

Anexo 5

Lista de comprobación para producción y publicación asistida por inteligencia artificial

La presente lista tiene por finalidad apoyar la fase de elaboración, revisión y publicación de contenidos cuando la inteligencia artificial haya intervenido de manera sustantiva en la redacción, estructuración, resumen, adaptación, análisis o transformación de información.

Su aplicación deberá ajustarse al nivel de riesgo del producto. No será necesario completar formalmente esta lista para cada corrección ortográfica o tarea menor. En cambio, podrá utilizarse como control efectivo para investigaciones, alertas, verificaciones, informes, contenidos jurídicos, materiales audiovisuales o publicaciones que puedan generar consecuencias relevantes.

La lógica general será:

> **PREPARAR → INSTRUIR → GENERAR → REVISAR → VERIFICAR → APROBAR → PUBLICAR → ARCHIVAR**

1. Confirmar que la información está lista para producción

Antes de utilizar IA para redactar un contenido sustantivo deberá comprobarse que la investigación haya alcanzado un grado suficiente de validación.

La persona responsable deberá preguntarse:

¿Los hechos centrales están verificados?

¿Las fuentes principales están identificadas?

¿Las contradicciones relevantes fueron revisadas?

¿Está claro qué información es confirmada y qué información permanece condicionada o pendiente?

¿Las citas que podrían utilizarse están comprobadas?

¿Las cifras relevantes pueden rastrearse hasta su fuente?

¿Existe información descartada que deba mantenerse fuera del proceso de redacción?

Si la IA recibe una mezcla de información validada, hipótesis y datos descartados sin una distinción clara, podrá incorporarlos indistintamente al texto.

2. Preparar el input de producción

El material proporcionado a la herramienta deberá ser suficiente para la tarea, pero no excesivo.

Cuando corresponda, deberá separarse en categorías como:

HECHOS VERIFICADOS

Información que puede utilizarse como base factual.

CITAS VERIFICADAS

Declaraciones cuya literalidad ha sido comprobada.

POSICIONES DE LAS PARTES

Versiones, denuncias, explicaciones u opiniones que deben mantenerse atribuidas.

CONTEXTO

Antecedentes útiles para comprender el asunto.

INFORMACIÓN CONDICIONADA

Datos que solo pueden utilizarse manteniendo expresamente su limitación.

RESTRICCIONES

Información que la herramienta no debe utilizar, inferir o completar.

La claridad del input reduce la posibilidad de que la IA trate una acusación como hecho o una hipótesis como conclusión.

3. Aplicar minimización antes de redactar

El hecho de que una información haya sido recopilada legítimamente no significa que toda ella deba proporcionarse a la herramienta de producción.

Antes de cargar documentos o expedientes deberá eliminarse aquello que no sea necesario para la tarea.

Esto puede comprender:

- nombres de fuentes protegidas;
- datos de contacto;
- números de identificación;
- anexos irrelevantes;
- información personal de terceros;
- documentos completos cuando únicamente se requiere un fragmento; o
- datos restringidos que no tengan relevancia para la redacción.

La fase de producción no deberá ampliar innecesariamente la exposición de información sensible.

4. Verificar la herramienta y el entorno

Antes de iniciar la producción deberá comprobarse que:

☐ se utiliza la cuenta institucional correspondiente;

☐ la herramienta está autorizada para el tipo de información;

☐ el proyecto o espacio tiene los permisos correctos;

☐ la información no se está trasladando a una cuenta personal;

☐ no existen integraciones innecesarias activadas para la tarea; y

☐ la memoria o repositorio utilizado corresponde al proyecto adecuado.

Cuando la producción sea de riesgo alto, deberá comprobarse también que el sistema no tenga acceso innecesario a otros proyectos o fuentes restringidas.

5. Construir una instrucción adecuada

El prompt deberá establecer con suficiente claridad qué espera Fundamedios del sistema.

Para productos relevantes podrá incluir:

Objetivo: qué debe producirse.

Audiencia: para quién se escribe.

Formato: nota, alerta, informe, resumen, guion, publicación, etc.

Fuentes permitidas: qué información puede utilizar.

Limitaciones: qué no debe incorporar.

Atribuciones: qué afirmaciones deben permanecer vinculadas a una fuente.

Incertidumbres: qué elementos no pueden presentarse como certeza.

Extensión y estilo: características editoriales requeridas.

Prohibición de completar: instrucción expresa para no inventar datos faltantes.

Tratamiento de información nueva: cualquier dato adicional deberá señalarse como pendiente de verificación.

6. Incluir restricciones cuando sean necesarias

En contenidos sensibles podrá resultar conveniente incorporar instrucciones negativas expresas.

Por ejemplo:

No inventes nombres, fechas, cifras, fuentes, enlaces ni citas.

No conviertas acusaciones en hechos.

No elimines expresiones de atribución.

No completes información ausente.

No identifiques fuentes protegidas.

No atribuyas intenciones sin evidencia.

No utilices información diferente de los materiales autorizados.

Si encuentras una contradicción, señálala en lugar de resolverla por inferencia.

Estas restricciones deberán adaptarse al producto y no utilizarse como una fórmula mecánica cuando no resulten pertinentes.

7. Tratar el primer resultado como borrador

El primer resultado generado por la IA deberá considerarse material de trabajo.

No deberá copiarse y publicarse directamente.

La persona responsable deberá leerlo íntegramente y evaluar:

qué añadió;

qué eliminó;

qué simplificó;

qué interpretó;

qué afirmó con mayor seguridad que las fuentes; y

qué pudo haber modificado respecto de los insumos.

Un texto fluido puede ocultar errores de fondo.

8. Comprobar que no existan hechos nuevos

Uno de los controles más importantes será identificar si el sistema incorporó información que no estaba en los insumos autorizados.

Cuando aparezca un hecho nuevo deberá preguntarse:

¿De dónde salió esto?

Si no puede vincularse con una fuente validada, deberá eliminarse o regresar a investigación.

La producción no deberá convertirse en una etapa silenciosa de adquisición de nuevos hechos.

9. Verificar afirmaciones centrales

Las afirmaciones sobre las cuales descansa el contenido deberán revisarse directamente contra sus fuentes.

No será suficiente que el texto “coincida en términos generales” con la investigación.

Deberá comprobarse especialmente:

- hechos principales;
- acusaciones;
- relaciones entre personas o instituciones;
- cifras centrales;
- decisiones judiciales o administrativas;
- antecedentes;
- afirmaciones que sustenten el titular; y
- conclusiones de verificaciones.

La pregunta práctica será:

¿Si esta afirmación fuera incorrecta, cambiaría sustancialmente el sentido de la publicación?

Si la respuesta es sí, deberá recibir revisión reforzada.

10. Revisar citas

Antes de publicar una cita textual deberá comprobarse:

☐ que existe;

☐ que corresponde a la persona indicada;

☐ que las palabras son correctas;

☐ que el contexto no cambia su sentido;

☐ que la traducción, cuando exista, es fiel; y

☐ que no fue reconstruida por la IA.

Cuando no pueda comprobarse la literalidad deberá utilizarse una paráfrasis correctamente atribuida.

11. Revisar cifras

Toda cifra material deberá comprobarse.

Especialmente:

monto;

moneda;

porcentaje;

período;

universo;

unidad;

fuente;

metodología;

fecha de actualización.

Si la IA realizó cálculos, deberá revisarse el resultado.

Cuando la cifra sea central para la publicación, deberá conservarse evidencia suficiente para reproducir el cálculo.

12. Revisar nombres, cargos y fechas

Antes de publicar deberán comprobarse los nombres propios, denominaciones institucionales, cargos y fechas relevantes.

Deberá prestarse atención a:

- homónimos;
- cargos actuales frente a cargos históricos;
- nombres oficiales de instituciones;
- fechas de nombramiento o salida;
- cambios de denominación;
- segundos apellidos;
- siglas; y
- cronologías.

La IA puede completar automáticamente estos elementos con información plausible pero incorrecta.

13. Revisar atribuciones

Las expresiones de atribución no deberán eliminarse por razones de estilo.

Deberá comprobarse si el texto distingue correctamente:

hechos verificados;

denuncias;

opiniones;

versiones de las partes;

afirmaciones de terceros;

interpretaciones; y

conclusiones propias de Fundamedios.

Una acusación atribuida no deberá convertirse en una afirmación categórica durante la reescritura.

14. Revisar situación jurídica y presunción de inocencia

Cuando el contenido involucre procesos penales, administrativos, disciplinarios o judiciales deberá comprobarse que la terminología corresponda al estado real del procedimiento.

No deberán utilizarse como equivalentes:

investigado; procesado; acusado; sentenciado; condenado; sancionado; denunciado.

La redacción deberá mantener las diferencias jurídicas correspondientes.

Cuando no exista una decisión definitiva, no deberá atribuirse responsabilidad como hecho consumado.

15. Revisar versiones y contraste

Cuando el producto requiera contraste, deberá comprobarse que la versión de la persona o institución involucrada haya sido adecuadamente considerada.

La revisión deberá preguntar:

¿Se solicitó su posición cuando correspondía?

¿Su respuesta fue presentada correctamente?

¿La IA eliminó elementos que modificaban sustancialmente su versión?

¿Se informó correctamente cuando no existió respuesta?

No deberá utilizarse IA para reducir una respuesta de manera que deje únicamente los fragmentos convenientes para la narrativa principal.

16. Revisar contexto

Una redacción resumida puede ser literalmente cierta y, al mismo tiempo, producir una impresión equivocada.

Deberá comprobarse si faltan antecedentes esenciales.

Particularmente:

fechas;

excepciones;

cambios posteriores;

limitaciones metodológicas;

alcance de una decisión;

universo de una cifra;

circunstancias de una declaración; o

cualquier elemento que cambie razonablemente la interpretación del hecho.

La brevedad no deberá obtenerse sacrificando contexto indispensable.

17. Revisar inferencias

La IA puede presentar conclusiones plausibles como si fueran hechos.

Deberá identificarse especialmente lenguaje relacionado con:

causalidad;

intenciones;

motivaciones;

coordinación;

responsabilidad;

relaciones personales;

conflictos de interés;

patrones de conducta; o

estados mentales.

Cuando una afirmación sea inferencial deberá comprobarse si existe evidencia suficiente para sostenerla.

Si no existe, deberá presentarse como hipótesis o eliminarse.

18. Revisar afirmaciones jurídicas

Cuando el texto incluya una conclusión jurídica, deberán comprobarse las fuentes normativas y jurisprudenciales relevantes.

La revisión deberá asegurarse de que:

la norma existe;

está vigente;

el artículo citado es correcto;

la sentencia existe;

la decisión realmente sostiene lo que se le atribuye;

no se confunden hechos con criterios jurídicos; y

no se presenta como pacífica una interpretación controvertida.

Cuando la cuestión sea compleja o de alto impacto deberá considerarse revisión jurídica especializada.

19. Revisar titulares y subtítulos

El titular deberá revisarse de forma independiente.

La IA puede producir titulares que excedan lo que permite afirmar el cuerpo del texto.

Deberá comprobarse que el titular:

☐ esté respaldado por la evidencia;
☐ no convierta una denuncia en un hecho;
☐ no elimine una condición esencial;
☐ no atribuya culpabilidad sin respaldo;
☐ no exagere una cifra;
☐ no generalice indebidamente; y
☐ no cree una impresión que el texto posteriormente tenga que desmentir.

El mismo control deberá aplicarse a subtítulos, sumarios y textos destacados.

20. Revisar publicaciones para redes sociales

Cuando la IA adapte un contenido para redes, deberá comprobarse cada pieza.

Un hilo, carrusel o publicación breve deberá conservar precisión incluso cuando sea visto separadamente del contenido completo.

No deberá suponerse que:

“el enlace explica el contexto”

si la propia publicación contiene una afirmación engañosa o categórica.

Las piezas de distribución deberán someterse a los mismos estándares factuales esenciales del contenido principal.

21. Revisar imágenes

Cuando se utilicen imágenes deberá determinarse su origen.

Fotografía o imagen documental

Debe conservar relación con el hecho representado.

Ilustración

Debe ser reconocible como tal cuando exista riesgo de confusión.

Imagen generada por IA

Deberá identificarse cuando pueda confundirse con material real.

Imagen real modificada sustancialmente

Deberá evaluarse si la modificación requiere disclosure y si continúa siendo adecuada para uso periodístico.

Nunca deberá utilizarse una imagen sintética como prueba de un acontecimiento real.

22. Revisar audio y video

Cuando exista contenido generado, doblado, clonado o alterado mediante IA deberá comprobarse si el público puede confundirlo con material auténtico.

Cuando corresponda deberá identificarse claramente su naturaleza.

No deberán utilizarse voces sintéticas o representaciones audiovisuales para atribuir a una persona palabras o acciones que no realizó.

Las recreaciones deberán mantenerse claramente separadas del material documental.

23. Protección de datos antes de publicar

Antes de la publicación deberá revisarse si el contenido incluye información personal innecesaria.

La pregunta deberá ser:

¿Necesita el público conocer este dato para comprender la historia?

Deberá prestarse especial atención a:

- direcciones;
- teléfonos;
- correos;
- identificaciones;
- información bancaria;
- datos de familiares;
- ubicación;
- datos sensibles;
- información de menores;
- información de víctimas; y
- otros elementos cuya publicación pueda generar riesgos.

Que un dato haya sido utilizado durante la investigación no significa que deba publicarse.

24. Protección de fuentes

Antes de publicar deberá comprobarse que el producto no revele directa o indirectamente información protegida sobre una fuente.

La revisión deberá alcanzar:

texto;

capturas;

documentos adjuntos;

metadatos;

fechas;

descripciones del cargo;

lugares;

fragmentos de conversaciones; y

cualquier combinación de elementos que pueda permitir identificarla.

En asuntos sensibles deberá realizarse una revisión específica de riesgo de identificación.

25. Revisar documentos adjuntos

Cuando una publicación permita descargar documentos, deberá revisarse el archivo completo antes de hacerlo público.

Deberá comprobarse que no contenga:

- información personal innecesaria;
- datos de fuentes;
- comentarios internos;
- control de cambios;
- versiones anteriores;
- hojas ocultas;
- metadatos sensibles;
- firmas o identificadores que deban protegerse; o
- información que no haya sido autorizada para publicación.

La publicación de un documento completo puede producir una exposición que no aparece en el cuerpo de la nota.

26. Realizar revisión adversarial cuando corresponda

En contenidos de riesgo medio o alto podrá realizarse una revisión adicional cuyo objetivo sea encontrar problemas y no mejorar el estilo.

Una instrucción posible será:

“Revise este borrador contra las fuentes proporcionadas. Identifique afirmaciones sin respaldo, cambios de atribución, citas no verificables, cifras inconsistentes, conclusiones que excedan la evidencia, omisiones de contexto, posibles riesgos de privacidad y cualquier elemento que no aparezca en los insumos.”

El resultado deberá entenderse como apoyo.

La ausencia de advertencias de la IA no sustituirá la revisión humana.

27. Segunda revisión humana

Los productos de mayor riesgo deberán contar, cuando corresponda, con una segunda revisión.

El revisor deberá examinar especialmente:

afirmaciones centrales;

fuentes;

acusaciones;

citas;

datos sensibles;

titular;

contexto; y

posibles consecuencias jurídicas o reputacionales.

Siempre que sea posible, el revisor deberá acceder directamente a las fuentes principales y no únicamente al borrador.

28. Revisión especializada

Antes de la publicación deberá determinarse si el producto requiere revisión adicional.

JURÍDICA

Cuando exista riesgo legal relevante.

PROTECCIÓN DE DATOS

Cuando existan tratamientos o publicaciones de información personal de alto riesgo.

SEGURIDAD

Cuando la publicación pueda comprometer sistemas, fuentes o información protegida.

TÉCNICA

Cuando la conclusión dependa de conocimientos especializados.

La IA podrá apoyar estas revisiones, pero no sustituirlas cuando sean necesarias.

29. Determinar si corresponde disclosure

Antes de publicar deberá revisarse si la intervención de inteligencia artificial resulta material para comprender el producto.

La divulgación podrá ser necesaria cuando exista:

- generación sustantiva de imágenes;
- audio o video sintético;
- modificación sustancial de material documental;
- análisis automatizado central para las conclusiones;
- procesamiento masivo relevante;

- metodología basada sustancialmente en IA; o
- cualquier otra intervención cuya ausencia de explicación pueda inducir a error sobre cómo se produjo el contenido.

No será necesario divulgar individualmente usos puramente instrumentales cuando la declaración institucional general resulte suficiente.

30. Revisar el disclosure

Cuando corresponda una divulgación específica deberá comprobarse que describa correctamente la función de la IA.

No deberá utilizarse una fórmula genérica que minimice una intervención relevante.

Tampoco deberá exagerarse una participación menor.

La transparencia deberá ser:

verdadera + específica + proporcional

31. Aprobar la versión exacta

La versión aprobada deberá ser la versión que se publique.

Después de la aprobación no deberá solicitarse a la IA una nueva “mejora final”, resumen, ajuste de tono o reformulación sustantiva sin volver a revisar el resultado.

Cuando exista una modificación posterior deberá determinarse si requiere nueva aprobación.

Esta regla será especialmente importante para:

titulares;

acusaciones;

cifras;

citas;

conclusiones; y

contenidos de alto riesgo.

32. Automatización de la publicación

La automatización podrá utilizarse para ejecutar técnicamente la publicación de un producto ya aprobado.

Deberá mantenerse la diferencia entre:

automatizar la publicación

y

automatizar la decisión de publicar.

La primera podrá ser admisible.

La segunda no deberá aplicarse a productos sustantivos cuando el Manual exige aprobación humana.

33. Control final inmediatamente antes de publicar

En productos relevantes deberá realizarse una comprobación rápida de la versión que efectivamente aparecerá al público.

CONTENIDO

¿Es la versión aprobada?

TITULAR

¿Coincide con la evidencia?

NOMBRES

¿Son correctos?

FECHAS

¿Son correctas?

CIFRAS

¿Son correctas?

CITAS

¿Fueron verificadas?

ENLACES

¿Funcionan y conducen a la fuente adecuada?

IMÁGENES

¿Corresponden al contenido?

DATOS PERSONALES

¿Se eliminó información innecesaria?

FUENTES

¿No existe riesgo de identificación?

DISCLOSURE

¿Se incorporó cuando corresponde?

ARCHIVOS

¿Los adjuntos fueron revisados?

34. Registro de publicación

Cuando la intervención de IA haya sido sustantiva y el nivel de riesgo lo justifique, deberá quedar registro de:

producto;

fecha;

responsable;

herramienta utilizada;

función realizada por IA;

principales fuentes;

revisión realizada;

aprobación; y

versión publicada.

El registro deberá mantenerse dentro del expediente correspondiente y no convertirse en un trámite independiente innecesario.

35. Conservar las fuentes centrales

Después de publicación deberán conservarse, durante el período correspondiente, los materiales necesarios para respaldar las afirmaciones centrales.

Esto podrá incluir:

- documentos;
- audios;
- videos;
- bases de datos;
- capturas;
- respuestas de fuentes;
- registros oficiales;
- verificaciones; y
- otros materiales relevantes.

Los resúmenes generados por IA no deberán sustituir a las fuentes originales cuando estas sean necesarias para defender o revisar posteriormente el producto.

36. Correcciones posteriores

Si después de publicación aparece un error, deberá determinarse si fue:

un error factual;

una omisión;

una atribución incorrecta;

una cifra equivocada;

una cita errónea;

un problema de contexto;

una información desactualizada; o

un error derivado del procesamiento con IA.

La corrección deberá alcanzar tanto al contenido público como, cuando corresponda, a las bases o memorias internas que originaron el error.

37. Productos derivados

Cuando la publicación original haya sido reutilizada para generar otros contenidos, deberá evaluarse si una corrección afecta también:

redes sociales;

boletines;

traducciones;

videos;

gráficos;

alertas;

presentaciones; o

nuevos productos producidos a partir del contenido inicial.

La corrección deberá interrumpir la cadena de reproducción del error.

38. Lista rápida de aprobación

Antes de publicar un contenido asistido sustantivamente por IA, podrá utilizarse la siguiente lista:

☐ Los hechos centrales están verificados.
☐ Las fuentes principales están identificadas.
☐ No existen datos nuevos generados por IA sin verificar.
☐ Las citas fueron comprobadas.
☐ Las cifras fueron comprobadas.
☐ Los nombres, cargos y fechas son correctos.
☐ Las acusaciones permanecen atribuidas.
☐ Se respetó la situación jurídica de las personas.
☐ Se incorporó el contraste cuando correspondía.
☐ El contexto es suficiente.
☐ No existen inferencias presentadas como hechos.
☐ El titular está plenamente respaldado.
☐ Las publicaciones para redes mantienen precisión.
☐ Los datos personales incluidos son necesarios.
☐ Las fuentes confidenciales permanecen protegidas.
☐ Los archivos adjuntos fueron revisados.
☐ El material audiovisual está correctamente identificado.
☐ Se realizó revisión especializada cuando era necesaria.

☐ Se incorporó disclosure cuando correspondía.

☐ La versión aprobada es exactamente la que se publicará.

☐ Una persona autorizada aprobó el contenido.

39. Regla final de producción y publicación

La inteligencia artificial podrá contribuir significativamente a la claridad, rapidez y capacidad de producción de Fundamedios, pero estas ventajas solo tendrán valor institucional si se conserva la relación entre el contenido y la evidencia que lo sustenta.

La última revisión no deberá preguntarse únicamente:

“¿Está bien escrito?”

Deberá preguntarse:

“¿Podemos sostener todo aquello que estamos a punto de publicar?”

La regla final será:

Ningún contenido asistido sustantivamente por inteligencia artificial deberá publicarse hasta que una persona responsable haya comprobado que los hechos materiales están sustentados, las incertidumbres se encuentran correctamente expresadas, la información protegida permanece protegida y la versión final refleja aquello que Fundamedios realmente puede afirmar.

Anexo 6

Formato de registro de incidentes relacionados con inteligencia artificial

El presente formato tiene por finalidad permitir que Fundamedios documente de manera proporcional los incidentes relevantes relacionados con la utilización de inteligencia artificial.

No deberá utilizarse para registrar cada error menor, reformulación defectuosa o respuesta insatisfactoria producida por una herramienta durante el trabajo cotidiano. Su uso estará destinado a situaciones en las que exista o pueda existir una afectación relevante a la calidad de la información, la seguridad, la confidencialidad, los datos personales, las fuentes, los sistemas institucionales, los derechos de terceros o los productos publicados.

El registro deberá permitir reconstruir razonablemente:

> **QUÉ OCURRIÓ → QUÉ INFORMACIÓN ESTUVO INVOLUCRADA → QUÉ HERRAMIENTA INTERVINO → QUÉ CONTROLES EXISTÍAN → QUÉ MEDIDAS SE ADOPTARON → QUÉ DEBE CORREGIRSE**

La documentación del incidente no tendrá por finalidad atribuir automáticamente responsabilidad a quien lo reporta. Su función principal será facilitar la contención, investigación, corrección y aprendizaje institucional.

1. Identificación básica del incidente

Código o número interno: _____

Fecha del reporte: _____

Fecha y hora aproximada del incidente: _____

Persona que reporta: _____

Cargo / función: _____

Proyecto / área: _____

Herramienta de IA involucrada: _____

Cuenta o entorno utilizado: _____

¿El incidente continúa activo? ☐ Sí ☐ No ☐ No se ha podido determinar

2. Descripción inicial

Describa brevemente qué ocurrió.

La descripción inicial deberá concentrarse en los hechos conocidos al momento del reporte. No será necesario establecer inmediatamente la causa ni determinar responsabilidades.

Cuando existan elementos todavía inciertos, deberá indicarse expresamente.

3. Tipo de incidente

Marque una o varias categorías cuando correspondan.

INFORMACIÓN Y CALIDAD EDITORIAL

☐ Información falsa generada por IA.

☐ Cita inexistente o incorrecta.

☐ Fuente o referencia inexistente.

☐ Cifra incorrecta.

☐ Error de identidad o confusión de personas.

☐ Información desactualizada.

☐ Inferencia presentada como hecho.

☐ Eliminación indebida de atribución o contexto.

☐ Información descartada reincorporada por el sistema.

☐ Otro error factual o editorial.

PROTECCIÓN DE DATOS

☐ Datos personales cargados en una herramienta no autorizada.

☐ Exposición o acceso indebido a datos personales.

☐ Datos sensibles involucrados.

☐ Conservación o reutilización indebida.

☐ Comunicación no autorizada de datos.

☐ Otro posible incidente de protección de datos.

CONFIDENCIALIDAD

☐ Información confidencial introducida en una herramienta no autorizada.

☐ Documento reservado expuesto.

☐ Investigación no publicada accesible a personas no autorizadas.

☐ Información institucional transferida a una cuenta personal.

☐ Otro incidente de confidencialidad.

FUENTES PERIODÍSTICAS

☐ Identidad de una fuente expuesta.

☐ Información que podría permitir inferir la identidad de una fuente.

☐ Material proporcionado por una fuente procesado fuera del entorno autorizado.

☐ Otro riesgo relacionado con protección de fuentes.

CIBERSEGURIDAD

☐ Cuenta comprometida.

☐ Credenciales expuestas.

☐ Dispositivo perdido o comprometido.

☐ Acceso no autorizado.

☐ Integración o conector utilizado indebidamente.

☐ Posible prompt injection.

☐ Ejecución de una acción no autorizada.

☐ Pérdida, modificación o eliminación de información.

☐ Otro incidente de seguridad.

PUBLICACIÓN

☐ Contenido incorrecto publicado.

☐ Contenido publicado sin aprobación.

☐ Versión distinta de la aprobada.

☐ Publicación automatizada no autorizada.

☐ Imagen, audio o video sintético incorrectamente identificado.

☐ Otro incidente relacionado con publicación.

4. Información afectada

Indique qué tipo de información estuvo involucrada:

☐ Pública

☐ Interna

☐ Confidencial

☐ Restringida

☐ Datos personales

☐ Datos sensibles

☐ Información de fuente confidencial

☐ Credenciales o información de seguridad

☐ Información jurídica protegida

☐ Investigación no publicada

☐ No se ha podido determinar

Descripción de la información afectada:

Deberá evitarse reproducir dentro de este formulario información especialmente sensible cuando no resulte necesario. Cuando corresponda, podrá utilizarse una referencia al expediente protegido en el que se encuentren los detalles.

5. Personas o sistemas potencialmente afectados

Indique, cuando sea posible:

Personas afectadas o potencialmente afectadas:

Proyectos o investigaciones involucrados:

Sistemas o repositorios relacionados:

Terceros o proveedores que podrían haber recibido información:

Cuando no exista certeza deberá indicarse que la información se encuentra pendiente de confirmación.

6. Cómo se detectó el incidente

☐ Detectado por la propia persona usuaria.

☐ Detectado durante revisión editorial.

☐ Detectado durante verificación.

☐ Detectado por otra persona de Fundamedios.

☐ Reportado por una fuente.

☐ Reportado por una persona mencionada en una publicación.

☐ Reportado por el público.

☐ Detectado por un sistema de seguridad.

☐ Notificado por el proveedor de la herramienta.

☐ Detectado durante auditoría o revisión interna.

☐ Otro.

Descripción:

7. Estado del producto relacionado

Cuando el incidente se encuentre vinculado con un contenido editorial, indique:

☐ No existía todavía un borrador.

☐ Se encontraba en investigación.

☐ Existía un borrador interno.

☐ Se encontraba en revisión.

☐ Estaba aprobado pero no publicado.

☐ Fue publicado.

☐ Fue distribuido también en redes sociales u otros canales.

☐ No aplica.

Producto relacionado: _____

Enlace o referencia interna: _____

8. Secuencia preliminar del incidente

Cuando resulte posible, reconstruya brevemente la secuencia:

1. INPUT

¿Qué información recibió la herramienta?

2. INSTRUCCIÓN

¿Qué se le solicitó?

3. OUTPUT

¿Qué produjo?

4. ACTUACIÓN HUMANA

¿Qué hizo posteriormente la persona usuaria?

5. REVISIÓN

¿Qué controles se realizaron?

6. RESULTADO

¿Cómo se materializó o detectó el incidente?

Esta sección podrá completarse posteriormente si al momento del reporte no existe información suficiente.

9. Medidas inmediatas adoptadas

Indique qué medidas se realizaron al detectar el incidente:

☐ Se detuvo el uso de la herramienta.
☐ Se suspendió una integración.
☐ Se restringió el acceso.
☐ Se cerraron sesiones.
☐ Se cambió o revocó una credencial.
☐ Se retiró información de la herramienta.
☐ Se preservaron registros.
☐ Se suspendió una publicación.
☐ Se retiró temporalmente un contenido.
☐ Se corrigió una publicación.
☐ Se notificó al responsable editorial.
☐ Se notificó al responsable tecnológico.
☐ Se solicitó revisión jurídica.
☐ Se activó el procedimiento de protección de datos.
☐ Se adoptaron medidas para proteger una fuente.
☐ No se adoptaron todavía medidas.

☐ Otra.

Detalle:

10. Evaluación preliminar de gravedad

Clasifique provisionalmente el incidente.

MENOR

Error contenido dentro del flujo ordinario y corregido antes de producir consecuencias relevantes.

☐

MODERADO

Incidente que requiere intervención adicional, afectó un proceso relevante o produjo una exposición limitada.

☐

GRAVE

Puede producir una afectación significativa a personas, información, productos, reputación, seguridad o derechos.

☐

CRÍTICO

Puede comprometer gravemente fuentes, seguridad física, información altamente protegida, sistemas esenciales o producir consecuencias difícilmente reversibles.

☐

Justificación de la clasificación:

La clasificación podrá modificarse posteriormente a medida que se conozca mayor información.

11. Evaluación del alcance

Una vez realizada la revisión inicial, deberá procurarse responder:

¿Qué información estuvo realmente comprometida?

¿Quién pudo acceder a ella?

¿Durante cuánto tiempo?

¿La información salió del entorno institucional?

☐ Sí ☐ No ☐ No determinado

¿El incidente alcanzó al público?

☐ Sí ☐ No ☐ No determinado

¿Existe riesgo de que continúe produciendo efectos?

☐ Sí ☐ No ☐ No determinado

¿Existen productos derivados que puedan estar afectados?

☐ Sí ☐ No ☐ No determinado

12. Evaluación de datos personales

Cuando exista información personal involucrada:

Categoría de datos: _____

Número aproximado de personas afectadas, cuando pueda determinarse:

¿Existen datos sensibles o de protección reforzada? ☐ Sí ☐ No ☐ No determinado

¿Puede existir riesgo para los derechos de los titulares? ☐ Sí ☐ No ☐ En evaluación

¿Se requiere análisis conforme a la normativa de protección de datos? ☐ Sí ☐ No ☐ En evaluación

Responsable de la revisión: _____

13. Evaluación de fuentes confidenciales

Cuando el incidente pueda afectar una fuente:

¿Se reveló directamente la identidad? ☐ Sí ☐ No ☐ No determinado

¿La combinación de datos puede permitir identificarla? ☐ Sí ☐ No ☐ No determinado

¿Existe riesgo para la seguridad o situación de la fuente? ☐ Sí ☐ No ☐ En evaluación

¿Se adoptaron medidas específicas de protección? ☐ Sí ☐ No ☐ No aplica

Detalle bajo acceso restringido: Referencia: _____

Los datos capaces de identificar a la fuente no deberán reproducirse innecesariamente dentro de un registro general de incidentes.

14. Evaluación editorial

Cuando el incidente haya afectado un producto publicado o en elaboración:

¿La información incorrecta era central para el producto? ☐ Sí ☐ No

¿Afectaba el titular? ☐ Sí ☐ No

¿Afectaba una acusación o señalamiento? ☐ Sí ☐ No

¿Afectaba una cita? ☐ Sí ☐ No

¿Afectaba una cifra relevante? ☐ Sí ☐ No

¿Modificaba sustancialmente la conclusión? ☐ Sí ☐ No

Medida editorial adoptada:

☐ Corrección interna antes de publicación.

☐ Actualización.

☐ Corrección pública.

☐ Aclaración.

☐ Retiro temporal.

☐ Retiro definitivo.

☐ No se requirió actuación pública.

☐ Pendiente de decisión.

15. Preservación de evidencia

Indique qué elementos fueron preservados:

☐ Conversación de IA relevante.

☐ Prompt.

☐ Output original.

☐ Archivos proporcionados.

☐ Fuentes utilizadas.

☐ Versiones del producto.

☐ Registros de acceso.

☐ Capturas.

☐ Configuración de la herramienta.

☐ Registros de aprobación.

☐ Comunicaciones relevantes.

☐ Otro.

Ubicación institucional de la evidencia:

La evidencia deberá almacenarse de acuerdo con su nivel de clasificación y acceso.

16. Análisis de causa

Una vez reconstruido el proceso deberá identificarse, en la medida razonablemente posible, la causa o causas que contribuyeron al incidente.

Podrán considerarse:

FUENTE

☐ La información original era incorrecta.

☐ La fuente fue interpretada fuera de contexto.

☐ Se utilizó una fuente desactualizada.

INTELIGENCIA ARTIFICIAL

☐ Alucinación.

☐ Error de extracción.

☐ Error de transcripción.

☐ Error de clasificación.

☐ Error de resumen.

☐ Inferencia incorrecta.

☐ Error de cálculo.

☐ Comportamiento inesperado del sistema.

INSTRUCCIÓN

☐ Prompt insuficiente.

☐ Instrucción ambigua.

☐ No se establecieron restricciones necesarias.

☐ Se proporcionó información incorrecta o incompleta.

PROCESO HUMANO

☐ No se verificó un dato que debía verificarse.

☐ Revisión insuficiente.

☐ Se utilizó una versión equivocada.

☐ Error de carga o selección de archivo.

☐ Uso de una cuenta o herramienta incorrecta.

ORGANIZACIÓN

☐ Herramienta inadecuada.

☐ Configuración insuficiente.

☐ Permisos excesivos.

☐ Falta de capacitación.

☐ Procedimiento poco claro.

☐ Falta de supervisión.

☐ Control inexistente o ineficaz.

SEGURIDAD

☐ Credencial comprometida.

☐ Integración insegura.

☐ Prompt injection.

☐ Acceso no autorizado.

☐ Vulnerabilidad o configuración técnica.

OTRA CAUSA

Más de una causa podrá contribuir al mismo incidente.

17. Controles que debían operar

Indique qué controles eran aplicables:

☐ Clasificación de información.

☐ Uso de cuenta institucional.

☐ Minimización.

☐ Verificación de fuentes.

☐ Revisión de citas.

☐ Revisión de cifras.

☐ Segunda revisión humana.

☐ Revisión editorial.

☐ Revisión jurídica.

☐ Control de acceso.

☐ Autenticación.

☐ Aprobación previa a publicación.

☐ Disclosure.

☐ Otro.

¿Los controles fueron aplicados?

☐ Sí, adecuadamente.

☐ Sí, parcialmente.

☐ No.

☐ No existía un control suficiente.

☐ No determinado.

18. Determinación de la falla

A partir del análisis anterior, identifique de manera preliminar:

☐ Error tecnológico detectable solo posteriormente.

☐ Error razonable dentro de un procedimiento adecuadamente aplicado.

☐ Falla de procedimiento.

☐ Falla de supervisión.

☐ Deficiencia institucional.

☐ Incumplimiento por falta de diligencia.

☐ Incumplimiento deliberado.

☐ Combinación de causas.

☐ No determinado.

Explicación:

Esta clasificación deberá realizarse después de reconstruir los hechos y no utilizarse como conclusión automática por el solo resultado del incidente.

19. Medidas correctivas

Indique las medidas adoptadas o recomendadas:

SOBRE LA INFORMACIÓN

☐ Corregir el dato.

☐ Eliminar información incorrecta.

☐ Actualizar repositorios.

☐ Retirar información de una memoria activa.

☐ Restringir acceso.

SOBRE LA HERRAMIENTA

☐ Cambiar configuración.

☐ Deshabilitar funcionalidad.

☐ Suspender herramienta.

☐ Revisar proveedor.

☐ Eliminar integración.

SOBRE EL PROCEDIMIENTO

☐ Modificar prompt institucional.

☐ Incorporar un nuevo control.

☐ Reforzar verificación.

☐ Exigir segunda revisión.

☐ Modificar clasificación de riesgo.

SOBRE USUARIOS

☐ Capacitación.

☐ Revisión de permisos.

☐ Supervisión adicional.

☐ Restricción de acceso.

☐ Aplicación de las medidas que correspondan conforme a otros instrumentos.

SOBRE SEGURIDAD

☐ Cambio de credenciales.

☐ MFA.

☐ Revisión de accesos.

☐ Segmentación de repositorios.

☐ Revisión técnica.

OTRAS

20. Corrección de memoria y bases de conocimiento

Cuando el incidente haya involucrado información incorrecta, deberá comprobarse si permanece dentro de sistemas capaces de reutilizarla.

Revise:

☐ Base de conocimiento general.

☐ Proyecto específico.

☐ Memoria del asistente.

☐ Expediente de investigación.

☐ Plantilla.

☐ Cronología.

☐ Base estructurada.

☐ Publicaciones anteriores.

☐ Productos derivados.

☐ No aplica.

Acciones realizadas:

La corrección del producto público no deberá considerarse suficiente cuando el origen interno del error continúe activo.

21. Comunicaciones realizadas

☐ Responsable editorial.

☐ Responsable tecnológico.

☐ Dirección.

☐ Responsable de protección de datos.

☐ Asesoría jurídica.

☐ Persona afectada.

☐ Fuente.

☐ Autoridad competente.

☐ Proveedor tecnológico.

☐ Público.

☐ No fue necesaria comunicación externa.

Detalle cuando corresponda:

22. Seguimiento

Responsable de seguimiento: _____

Medidas pendientes: _____

Fecha prevista de revisión: _____

¿El incidente puede considerarse contenido? ☐ Sí ☐ No

¿El incidente puede considerarse cerrado? ☐ Sí ☐ No

23. Cierre

Un incidente podrá cerrarse cuando razonablemente:

se haya contenido el problema;

se conozca suficientemente su alcance;

se hayan adoptado las correcciones necesarias;

se hayan cumplido las obligaciones legales o institucionales pertinentes;

se haya corregido la información dentro de los repositorios afectados; y

se hayan definido las medidas necesarias para reducir su repetición.

El cierre no impedirá realizar posteriormente análisis adicionales si aparece nueva información.

24. Lección institucional

Todo incidente relevante deberá dejar una conclusión útil para el sistema.

¿Qué aprendió Fundamedios de este incidente?

¿Qué deberá hacerse de forma diferente en el futuro?

¿Debe modificarse el Manual o alguno de sus anexos?

☐ Sí ☐ No ☐ Evaluar posteriormente

¿Debe modificarse una plantilla de prompt?

☐ Sí ☐ No

¿Debe modificarse la clasificación de riesgo de esta actividad?

☐ Sí ☐ No

¿Se requiere capacitación adicional?

☐ Sí ☐ No

25. Regla final

El registro de incidentes deberá utilizarse como herramienta de **control, corrección y aprendizaje**, no como sustituto de un análisis de responsabilidad ni como mecanismo para atribuir culpa automáticamente a quien reporta un problema.

La prioridad será identificar el punto en el que falló el sistema y corregirlo.

La regla final será:

Todo incidente relevante relacionado con inteligencia artificial deberá dejar una respuesta verificable a tres preguntas: qué ocurrió, qué se hizo para corregirlo y qué cambió para reducir la posibilidad de que vuelva a ocurrir.

Anexo 7

Catálogo de herramientas autorizadas para uso institucional

El presente Catálogo tiene por finalidad identificar las herramientas de inteligencia artificial que Fundamedios autoriza para actividades institucionales, así como establecer las condiciones, límites y controles aplicables a cada una.

El Catálogo deberá mantenerse actualizado y podrá modificarse sin necesidad de reformar íntegramente el Manual, siempre que las nuevas autorizaciones o restricciones respeten los principios y controles establecidos en este.

La inclusión de una herramienta en el Catálogo no significará que pueda utilizarse para cualquier finalidad ni con cualquier tipo de información.

Una misma plataforma podrá encontrarse autorizada para determinadas actividades y restringida para otras.

La evaluación deberá considerar:

> **QUÉ HERRAMIENTA ES → PARA QUÉ SE UTILIZA → QUÉ INFORMACIÓN PUEDE RECIBIR → QUÉ FUNCIONES TIENE → QUÉ RIESGOS PRESENTA → QUÉ CONTROLES REQUIERE**

1. Identificación de la herramienta

Para cada herramienta deberá registrarse, al menos:

Nombre de la herramienta: _____

Proveedor: _____

Sitio o servicio institucional: _____

Tipo de licencia: ☐ Institucional ☐ Empresarial ☐ Profesional ☐ Gratuita ☐ Otra

Fecha de incorporación: _____

Última revisión institucional: _____

Responsable de administración: _____

Área principal de utilización: _____

2. Estado de autorización

Cada herramienta deberá encontrarse dentro de una de las siguientes categorías:

AUTORIZADA

Puede utilizarse institucionalmente para las finalidades y niveles de información expresamente señalados.

AUTORIZADA CON RESTRICCIONES

Puede utilizarse únicamente bajo determinadas condiciones, por usuarios específicos, para ciertos proyectos o con determinados tipos de información.

EN EVALUACIÓN

Puede encontrarse bajo análisis técnico, jurídico o institucional. No deberá utilizarse con información institucional protegida hasta que exista una decisión de autorización.

SUSPENDIDA

Su utilización institucional ha sido temporalmente detenida mientras se revisa un incidente, cambio de condiciones, problema de seguridad o cualquier otra circunstancia relevante.

NO AUTORIZADA

No deberá utilizarse para actividades institucionales comprendidas por el Manual.

Estado actual: _____

3. Finalidades autorizadas

Indique para qué actividades puede utilizarse la herramienta.

☐ Búsqueda de información pública
☐ Investigación asistida
☐ Análisis documental
☐ Resumen
☐ Traducción
☐ Transcripción
☐ Extracción de datos
☐ Clasificación de documentos
☐ Comparación de fuentes
☐ Construcción de cronologías
☐ Análisis de bases de datos
☐ Apoyo para fact-checking
☐ Elaboración de borradores
☐ Corrección de estilo
☐ Adaptación para redes sociales
☐ Generación de titulares
☐ Revisión adversarial
☐ Revisión de consistencia
☐ Generación de imágenes
☐ Edición de imágenes
☐ Generación de audio
☐ Generación de video
☐ Automatizaciones
☐ Agentes
☐ Integraciones con sistemas externos
☐ Otra: _____

4. Tipo de información permitida

Indique qué nivel de información puede procesar la herramienta.

Tipo de información	Permitido	Permitido con restricciones	No permitido
Pública	☐	☐	☐
Interna	☐	☐	☐
Confidencial	☐	☐	☐
Restringida	☐	☐	☐
Datos personales	☐	☐	☐
Datos sensibles	☐	☐	☐
Información de fuentes protegidas	☐	☐	☐
Información jurídica interna	☐	☐	☐
Credenciales / secretos	☐	☐	☐

Observaciones o condiciones especiales:

5. Uso desde cuentas institucionales

Indique cómo debe accederse a la herramienta.

☐ Solo mediante cuenta institucional.

☐ Puede utilizarse con cuenta institucional individual.

☐ Puede utilizarse dentro de un espacio de equipo.

☐ Puede utilizarse únicamente mediante cuenta administrada.

☐ El uso personal no se encuentra permitido para información institucional.

☐ Requiere acceso específico por proyecto.

☐ Otra condición: _____

Cuando exista un entorno institucional disponible, deberá evitarse procesar información de Fundamedios mediante cuentas personales.

6. Memoria e historial

Indique las capacidades de conservación de la herramienta.

¿Conserva historial de conversaciones? ☐ Sí ☐ No ☐ Configurable ☐ No determinado

¿Dispone de memoria persistente? ☐ Sí ☐ No ☐ Configurable ☐ No determinado

¿La memoria institucional se encuentra habilitada? ☐ Sí ☐ No ☐ Parcialmente

¿Puede deshabilitarse el historial? ☐ Sí ☐ No ☐ No determinado

¿Puede eliminarse información posteriormente? ☐ Sí ☐ Parcialmente ☐ No ☐ No determinado

Condiciones relevantes:

7. Utilización de información para entrenamiento o mejora

Deberá registrarse si la información proporcionada a la plataforma puede ser utilizada por el proveedor para entrenamiento, mejora de modelos u otras finalidades.

☐ No utiliza información institucional para entrenamiento.

☐ Puede utilizarla si determinada configuración se encuentra activa.

☐ Puede utilizarla según sus términos generales.

☐ No se ha podido determinar suficientemente.

☐ Otra condición.

Configuración institucional aplicada:

Cuando exista la posibilidad de desactivar la utilización de información institucional para entrenamiento o mejora general, deberá preferirse esa configuración salvo que exista una razón institucional diferente debidamente evaluada.

8. Conservación de datos por el proveedor

Tiempo aproximado de conservación: _____

¿Existe eliminación administrativa? ☐ Sí ☐ No ☐ Parcial

¿Existen copias de respaldo del proveedor? ☐ Sí ☐ No ☐ No determinado

¿Existe un procedimiento para solicitar eliminación? ☐ Sí ☐ No ☐ No determinado

Observaciones:

9. Localización y transferencias de información

Cuando resulte relevante para protección de datos y seguridad deberá registrarse:

País o jurisdicción principal del proveedor: _____

¿Puede procesar información en otras jurisdicciones? ☐ Sí ☐ No ☐ No determinado

¿Utiliza subencargados o subproveedores? ☐ Sí

☐ No ☐ No determinado

¿Existe documentación contractual sobre transferencias? ☐ Sí ☐ No ☐ No aplica

Observaciones jurídicas o de protección de datos:

10. Seguridad

Deberá evaluarse si la herramienta dispone de:

☐ Autenticación multifactor
☐ Administración centralizada de usuarios
☐ Roles y permisos
☐ Inicio de sesión institucional
☐ Registro de actividad
☐ Restricción por proyectos
☐ Cifrado de información
☐ Controles de compartición
☐ Exportación de datos
☐ Eliminación administrativa
☐ Revocación de sesiones
☐ Gestión de integraciones
☐ Otras medidas relevantes

Observaciones de seguridad:

11. Integraciones y conectores

Indique si la herramienta puede conectarse con:

☐ Correo electrónico
☐ Almacenamiento institucional
☐ Google Drive
☐ Microsoft 365
☐ Bases de datos
☐ Sistemas editoriales
☐ Redes sociales
☐ Calendarios
☐ Navegadores
☐ APIs
☐ Aplicaciones de terceros
☐ Otros sistemas

Las integraciones deberán autorizarse separadamente cuando amplíen sustancialmente el acceso de la herramienta.

Integraciones autorizadas:

Integraciones expresamente prohibidas:

12. Capacidad de actuar

Indique qué nivel de autonomía posee la herramienta.

NIVEL 1 — ASISTENCIAL

Solo genera respuestas o recomendaciones.

☐

NIVEL 2 — INTERACTIVO

Puede consultar información mediante conectores o herramientas.

☐

NIVEL 3 — SEMIAUTOMATIZADO

Puede preparar acciones que posteriormente requieren confirmación.

☐

NIVEL 4 — AUTÓNOMO

Puede ejecutar acciones externas sin confirmación inmediata.

☐

Cuando existan capacidades de nivel 3 o 4 deberán describirse las restricciones y controles aplicables.

Acciones permitidas:

Acciones que requieren aprobación humana:

Acciones prohibidas:

---

13. Generación audiovisual

Cuando la herramienta permita producir o modificar imágenes, audio o video deberá registrarse:

☐ Genera imágenes
☐ Modifica imágenes
☐ Genera voz sintética
☐ Clona voces
☐ Genera video
☐ Modifica video
☐ Permite representaciones realistas de personas
☐ No aplica

Condiciones de uso:

Todo material sintético destinado al público deberá respetar las reglas de transparencia y autenticidad establecidas en el Manual.

14. Riesgos conocidos

Identifique los principales riesgos asociados con la herramienta.

Podrán incluir:

- alucinaciones;
- referencias inexistentes;
- errores jurídicos;
- errores de extracción;
- problemas de privacidad;
- memoria no controlada;
- acceso demasiado amplio;
- vulnerabilidad a prompt injection;
- dificultades de eliminación;
- falta de trazabilidad;
- generación audiovisual realista;
- integración con terceros;
- sesgos;
- desactualización;
- imposibilidad de controlar completamente el tratamiento de datos.

Riesgos específicos identificados:

15. Restricciones institucionales

Indique las reglas especiales que deberán observar los usuarios.

Ejemplos:

No cargar fuentes confidenciales.

No utilizar para expedientes jurídicos internos.

No habilitar memoria en proyectos sensibles.

No permitir acceso al correo institucional.

Verificar manualmente todas las referencias jurídicas.

No utilizar contenido audiovisual generado como evidencia.

No procesar datos personales a gran escala.

Restricciones aplicables:

16. Nivel de riesgo institucional de la herramienta

Indique el nivel general asignado:

☐ Bajo

☐ Medio

☐ Alto

☐ Crítico para determinadas funciones

La clasificación general no sustituirá la evaluación del riesgo de cada actividad.

Una herramienta de riesgo ordinario puede utilizarse dentro de un proyecto de riesgo alto.

17. Usuarios o áreas autorizadas

☐ Toda la organización

☐ Equipo editorial

☐ Ecuador Chequea

☐ Liberties 360

☐ Equipo de investigación

☐ Administración

☐ Equipo jurídico

☐ Tecnología

☐ Usuarios nominados

Usuarios / áreas específicas:

18. Capacitación requerida

☐ No requiere capacitación adicional.

☐ Capacitación institucional general.

☐ Capacitación específica sobre la herramienta.

☐ Capacitación sobre protección de datos.

☐ Capacitación sobre seguridad.

☐ Capacitación editorial.

☐ Capacitación técnica.

Requisitos antes de conceder acceso:

19. Incidentes registrados

¿Existen incidentes relevantes vinculados con la herramienta? ☐ Sí ☐ No

Referencias internas:

Los incidentes deberán considerarse al revisar la continuidad de la autorización.

20. Revisión periódica

La herramienta deberá revisarse nuevamente cuando:

cambien sus términos;

modifique su política de privacidad;

incorpore nuevas funciones;

habilite memoria;

incorpore agentes o automatizaciones;

cambie la forma en que utiliza información para entrenamiento;

se produzca un incidente significativo;

aparezcan riesgos nuevos; o

transcurra el período institucional establecido para revisión.

Próxima fecha de evaluación: _____

21. Decisión institucional

Después de la evaluación:

☐ AUTORIZADA
☐ AUTORIZADA CON RESTRICCIONES
☐ EN EVALUACIÓN
☐ SUSPENDIDA
☐ NO AUTORIZADA

Condiciones de la decisión:

Responsable de la evaluación: _____

Fecha: _____

Aprobación: _____

22. Catálogo resumido

Fundamedios podrá mantener adicionalmente una tabla general para consulta rápida:

Herramienta	Estado	Uso principal	Informaci ón máxima permitida	Usuarios	Revisión
_____	_____	_____	_____	_____	_____
_____	_____	_____	_____	_____	_____
_____	_____	_____	_____	_____	_____
_____	_____	_____	_____	_____	_____

Esta tabla permitirá a las personas conocer rápidamente qué herramientas se encuentran habilitadas sin necesidad de consultar toda la ficha técnica para cada uso ordinario.

23. Regla final

El Catálogo deberá responder a una pregunta práctica antes de que una persona utilice una nueva herramienta con información institucional:

¿Fundamedios conoce esta herramienta y ha definido para qué podemos utilizarla?

Cuando la respuesta sea afirmativa, deberán respetarse las condiciones establecidas.

Cuando sea negativa y la actividad involucre información institucional protegida, la herramienta deberá ser evaluada antes de utilizarse.

La regla final será:

Ninguna herramienta será considerada segura o apropiada para Fundamedios únicamente porque sea conocida, popular o técnicamente capaz. Su utilización institucional dependerá de una evaluación proporcional de sus funciones, tratamiento de información, riesgos y controles disponibles.

Anexo 8

Biblioteca inicial de prompts institucionales

La presente Biblioteca contiene modelos de instrucciones para algunas de las actividades recurrentes en las que Fundamedios, Ecuador Chequea, Liberties 360 y las demás iniciativas, proyectos y plataformas de la organización podrán utilizar inteligencia artificial.

Los prompts incluidos en este Anexo no deberán entenderse como fórmulas rígidas. Constituyen **estructuras de referencia** que podrán adaptarse a las características de cada investigación, producto, herramienta y nivel de riesgo.

Su utilización tampoco sustituirá las obligaciones de clasificación de información, protección de datos, verificación, revisión humana, trazabilidad y aprobación previstas en el Manual.

La lógica institucional será:

UN BUEN PROMPT DELIMITA LA TAREA; UNA BUENA FUENTE SUSTENTA EL RESULTADO; LA VERIFICACIÓN HUMANA DETERMINA SI PUEDE UTILIZARSE.

---

1. Estructura básica de un prompt institucional

No todas las instrucciones necesitarán contener los mismos elementos. Para trabajos sustantivos, sin embargo, será conveniente construir el prompt considerando cinco componentes:

CONTEXTO

Explicar a la herramienta qué material está recibiendo y para qué actividad se utilizará.

TAREA

Indicar con precisión qué debe hacer.

FUENTES

Determinar qué información puede utilizar y, cuando sea necesario, qué fuentes deberá priorizar.

RESTRICCIONES

Establecer aquello que no debe hacer, inferir, modificar o completar.

RESULTADO

Indicar cómo deberá presentar la información para facilitar su revisión.

Una estructura posible será:

Contexto: Estoy realizando una investigación sobre [ASUNTO].

Tarea: Analiza la documentación proporcionada para identificar [OBJETIVO].

Fuentes: Utiliza exclusivamente [FUENTES O DOCUMENTOS AUTORIZADOS].

Restricciones: No inventes información, no completes datos faltantes, no conviertas hipótesis en hechos y señala expresamente cualquier información que no pueda verificarse.

Resultado: Presenta los hallazgos indicando para cada uno la fuente, ubicación, nivel de certeza y verificaciones pendientes.

Esta estructura deberá adaptarse a cada caso.

2. Prompt para diseñar una estrategia de investigación

Este modelo podrá utilizarse antes de iniciar una investigación para identificar fuentes, documentos y líneas de búsqueda.

Estoy investigando **[DESCRIBIR ASUNTO]**.

El objetivo de la investigación es determinar **[PREGUNTA PRINCIPAL]**.

Diseña una estrategia de investigación sin asumir que la hipótesis inicial es correcta.

Identifica:

1. preguntas que deben responderse;
2. fuentes primarias que deberían buscarse;
3. registros públicos relevantes;
4. documentos oficiales potencialmente útiles;
5. fuentes secundarias que puedan orientar la búsqueda;
6. información necesaria para corroborar o descartar la hipótesis;
7. posibles explicaciones alternativas;
8. vacíos de información que probablemente deban resolverse mediante entrevistas o solicitudes de información.

No presentes como hechos datos que no hayan sido proporcionados. Si mencionas posibles fuentes o documentos específicos cuya existencia no puedes comprobar, identifícalos expresamente como sugerencias de búsqueda.

El resultado deberá entenderse como **plan de trabajo**, no como evidencia.

3. Prompt para búsqueda de información pública

Investiga **[ASUNTO]** durante el período comprendido entre **[FECHA]** y **[FECHA]**.

Prioriza fuentes primarias, documentos oficiales, registros públicos, publicaciones institucionales y fuentes periodísticas reconocidas.

Para cada hallazgo relevante proporciona:

hecho o información encontrada;

fuente;

fecha;

enlace o referencia;

fragmento o ubicación que permita comprobarlo;

si se trata de una fuente primaria o secundaria;

cualquier limitación o contradicción detectada.

No completes información ausente. Si no puedes encontrar evidencia suficiente sobre un punto, indica **“NO VERIFICADO”**.

4. Prompt para construir una cronología

Utiliza únicamente los documentos y fuentes proporcionados.

Construye una cronología de los acontecimientos relacionados con **[ASUNTO]**.

Organízala mediante:

> **FECHA | ACONTECIMIENTO | FUENTE | UBICACIÓN EN LA FUENTE | ESTADO DE VERIFICACIÓN**

Distingue entre:

HECHO VERIFICADO

AFIRMACIÓN ATRIBUIDA

INFORMACIÓN PENDIENTE

INFERENCIA

Si dos fuentes proporcionan fechas o versiones diferentes, no elijas una automáticamente. Registra la contradicción.

No incorpores acontecimientos que no aparezcan en las fuentes proporcionadas.

5. Prompt para análisis de un documento

Analiza el documento proporcionado con el objetivo de identificar información relacionada con **[OBJETIVO]**.

Para cada hallazgo indica:

- contenido relevante;
- página, sección o ubicación;
- significado dentro del contexto del documento;
- relación con la pregunta de investigación;
- cualquier limitación de interpretación.

Distingue claramente entre lo que **dice expresamente el documento** y cualquier **inferencia posible**.

No atribuyas al documento conclusiones que no se encuentren sustentadas por su texto.

6. Prompt para comparar documentos

Compara los documentos **[A]**, **[B]** y **[C]** respecto de **[ASUNTO]**.

Identifica:

coincidencias;

contradicciones;

información que aparece únicamente en una fuente;

cambios entre versiones;

omisiones relevantes;

fechas y cifras diferentes; y

puntos que requieren verificación adicional.

Para cada conclusión indica exactamente qué documento la sustenta.

No resuelvas contradicciones mediante inferencia cuando las fuentes no permitan hacerlo.

7. Prompt para extracción estructurada de información

Cuando deban analizarse numerosos documentos podrá utilizarse una estructura uniforme.

Extrae de cada documento únicamente los siguientes campos:

Documento: Fecha: Autor / institución: Persona mencionada: Cargo: Hecho relevante: Monto o cifra: Ubicación dentro del documento: Fuente: Observaciones:

Si un campo no aparece expresamente, escribe **"NO CONSTA"**.

No completes campos mediante conocimiento externo ni inferencias.

Esta última instrucción será especialmente importante para evitar que la IA "complete" bases de datos con información plausible.

8. Prompt para identificar contradicciones

Analiza los materiales proporcionados exclusivamente para identificar contradicciones.

Busca diferencias relacionadas con:

fechas; cifras; nombres; cargos; secuencia de acontecimientos; versiones de los involucrados; documentos; explicaciones causales; y cualquier otro elemento factual relevante.

Para cada contradicción presenta:

VERSIÓN A

FUENTE A

VERSIÓN B

FUENTE B

POSIBLE EXPLICACIÓN, únicamente si existe evidencia

VERIFICACIÓN ADICIONAL NECESARIA

No determines cuál versión es correcta cuando la documentación no permita hacerlo.

9. Prompt para buscar evidencia contraria a una hipótesis

Este prompt deberá utilizarse para reducir el riesgo de sesgo de confirmación.

La hipótesis preliminar de investigación es:

[HIPÓTESIS]

No intentes confirmarla.

Realiza una revisión adversarial de los materiales proporcionados e identifica:

1. evidencia que contradiga la hipótesis;
2. hechos que admitan una explicación diferente;
3. supuestos no demostrados;
4. información faltante;
5. posibles errores de causalidad;
6. coincidencias que podrían estar siendo interpretadas como relaciones;
7. información que favorecería a la persona o institución investigada;
8. verificaciones necesarias antes de sostener la hipótesis.

No construyas artificialmente argumentos contrarios. Señala únicamente aquellos que tengan sustento razonable.

10. Prompt para análisis de relaciones

Analiza la información proporcionada para identificar posibles relaciones entre **[PERSONAS / EMPRESAS / INSTITUCIONES]**.

Clasifica cada relación como:

DOCUMENTADA

INDICIO

COINCIDENCIA

INFERENCIA

NO DETERMINADA

Para cada relación indica la evidencia que la sustenta.

No interpretes coincidencias de nombres, direcciones, fechas, eventos, accionistas o contactos como prueba automática de coordinación, control, responsabilidad o actividad ilícita.

11. Prompt para análisis de una base de datos

Analiza la base proporcionada para identificar **[OBJETIVO]**.

Antes de presentar conclusiones:

1. describe qué campos contiene;
2. identifica datos faltantes;
3. detecta duplicados;
4. identifica inconsistencias;
5. señala posibles problemas de calidad;
6. explica cualquier transformación necesaria.

Después realiza el análisis solicitado.

Separa claramente:

RESULTADOS DIRECTAMENTE OBSERVABLES EN LOS DATOS

de

INFERENCIAS O INTERPRETACIONES.

No atribuyas causalidad únicamente por correlación.

12. Prompt para resumen fiel

Resume el documento proporcionado.

El resumen deberá conservar:

tesis principal;

hechos relevantes;

cifras;

condiciones;

excepciones;

conclusiones; y

limitaciones importantes.

No incorpores información externa.

No elimines condiciones que cambien el significado de una afirmación.

Si el documento contiene posiciones contrapuestas, conserva esa diferencia.

13. Prompt para revisión de una cita

Verifica si la siguiente cita:

"[CITA]"

aparece literalmente en el material proporcionado.

Si aparece, indica su ubicación exacta y el contexto inmediato.

Si la idea aparece pero con palabras diferentes, responde:

PARÁFRASIS, NO CITA TEXTUAL

y proporciona el texto real.

Si no aparece, responde:

CITA NO LOCALIZADA.

No reconstruyas ni completes la cita.

14. Prompt para análisis jurídico preliminar

Analiza jurídicamente **[ASUNTO]** conforme al ordenamiento jurídico de **[JURISDICCIÓN]**.

Identifica:

normativa aplicable;

artículos relevantes;

jurisprudencia pertinente;

posibles interpretaciones;

argumentos a favor y en contra;

aspectos que requieran verificación adicional.

Para cada norma o sentencia proporciona una referencia que permita comprobar su existencia.

No inventes números de artículos, causas, sentencias ni citas.

Si no puedes verificar una referencia, identifícala expresamente como **PENDIENTE DE VERIFICACIÓN**.

El resultado será utilizado como apoyo de investigación y deberá ser revisado contra las fuentes jurídicas originales.

15. Prompt para preparar un borrador periodístico

Utiliza exclusivamente los hechos y fuentes proporcionados para preparar un primer borrador de **[TIPO DE PRODUCTO]**.

Público objetivo: **[AUDIENCIA]**.

Extensión aproximada: **[EXTENSIÓN]**.

Mantén una redacción clara, precisa y proporcional a la evidencia.

Reglas:

- no añadas hechos externos;
- no inventes citas;
- no completes información faltante;
- mantén atribuidas las denuncias y afirmaciones de terceros;
- no conviertas hipótesis en hechos;
- conserva las incertidumbres relevantes;
- no atribuyas intenciones sin evidencia;
- no identifiques fuentes protegidas;
- no modifiques cifras.

Si detectas que falta información necesaria para redactar responsablemente una afirmación, coloca:

[VERIFICACIÓN PENDIENTE: …]
en lugar de completar el vacío.

16. Prompt para una alerta

Prepare un borrador de alerta institucional sobre **[HECHO]** utilizando exclusivamente la información validada proporcionada.

La alerta deberá identificar con claridad:

qué ocurrió;

a quién afecta;

cuándo ocurrió;

dónde ocurrió, cuando sea relevante;

fuente de la información;

contexto necesario; y

situación actual conocida.

No añada información no verificada.

Mantenga las acusaciones atribuidas a su fuente.

Cuando un elemento no haya podido determinarse, no lo complete por inferencia.

17. Prompt para fact-checking

Analiza la siguiente afirmación:

[AFIRMACIÓN]

No determines inicialmente si es verdadera o falsa.

Primero:

1. identifica exactamente qué afirmaciones verificables contiene;

2. determina qué evidencia sería necesaria para comprobar cada una;

3. identifica las fuentes primarias pertinentes;

4. separa hechos de opiniones o predicciones;

5. identifica el período temporal relevante.

Después analiza la evidencia proporcionada.

Para cada elemento indica:

> **VERIFICADO / FALSO / IMPRECISO / ENGAÑOSO POR CONTEXTO / NO DETERMINADO**

según la metodología institucional aplicable.

Explica la evidencia de cada conclusión y no utilices como sustento una afirmación generada por otra IA sin fuente independiente.

18. Prompt para detectar vacíos antes de redactar

Antes de redactar, revisa los materiales proporcionados y determina qué información falta para producir responsablemente un contenido sobre **[ASUNTO]**.

No redactes todavía.

Identifica únicamente:

hechos centrales ya comprobados;

hechos centrales todavía no comprobados;

contradicciones;

versiones que faltan;

citas pendientes de verificación;

cifras pendientes;

documentos que deberían localizarse;

posibles riesgos de privacidad;

posibles riesgos jurídicos; y

preguntas que deberían resolverse antes de publicación.

Este prompt podrá funcionar como control intermedio entre investigación y producción.

19. Prompt de revisión factual del borrador

Compara el borrador con las fuentes proporcionadas.

No mejores el estilo.

Realiza exclusivamente una auditoría factual.

Clasifica cada afirmación material como:

RESPALDADA

PARCIALMENTE RESPALDADA

NO RESPALDADA

CONTRADICHA POR UNA FUENTE

REQUIERE CONTEXTO

Para cada problema indica:

texto del borrador;

fuente correspondiente;

problema detectado;

corrección necesaria.

No asumas que una afirmación es correcta únicamente porque aparece redactada de manera convincente.

20. Prompt de revisión adversarial previa a publicación

Actúa como revisor crítico de este borrador.

Tu objetivo no es mejorar su redacción, sino encontrar razones por las cuales **no debería publicarse todavía**.

Busca específicamente:

1. hechos sin fuente;
2. acusaciones presentadas como hechos;
3. citas no verificadas;
4. cifras inconsistentes;
5. errores de identidad;
6. problemas cronológicos;
7. inferencias presentadas como certezas;
8. falta de contexto;
9. versiones relevantes omitidas;
10. riesgos de protección de datos;
11. posible identificación de fuentes;
12. afirmaciones jurídicas no sustentadas;
13. problemas entre titular y evidencia;
14. afirmaciones que exceden lo demostrado.

Para cada problema indica su gravedad:

> **BAJA / MEDIA / ALTA / CRÍTICA**

y qué debería verificarse antes de publicar.

21. Prompt para revisión de titular

Evalúa los siguientes titulares respecto del contenido y las fuentes proporcionadas:

[TITULARES]

Para cada uno determina:

¿Está completamente respaldado?

¿Convierte una acusación en hecho?

¿Exagera una cifra o conclusión?

¿Elimina una condición necesaria?

¿Puede producir una interpretación distinta de la evidencia?

¿Es más categórico que el cuerpo de la nota?

No propongas un titular más atractivo hasta terminar la revisión factual.

22. Prompt para adaptar contenido a redes sociales

Adapta el contenido aprobado proporcionado para **[PLATAFORMA]**.

No añadas hechos.

No modifiques cifras.

No elimines atribuciones necesarias.

No conviertas una afirmación condicionada en categórica.

Cada pieza deberá poder entenderse correctamente aunque sea vista fuera del contexto de la publicación completa.

Cuando la limitación de espacio impida mantener la precisión, prioriza exactitud sobre impacto.

23. Prompt para revisión de datos personales

Revisa este borrador exclusivamente desde la perspectiva de protección de datos y privacidad.

Identifica:

datos personales incluidos;

datos sensibles;

información de menores;

datos de víctimas;

direcciones o ubicaciones;

datos de contacto;

identificadores;

información cuya combinación pueda identificar a una persona;

datos que no parezcan necesarios para la finalidad de la publicación.

No elimines automáticamente información. Indica qué elementos requieren una decisión humana sobre necesidad, proporcionalidad o interés público.

24. Prompt para revisión de protección de fuentes

Revisa el contenido con el único objetivo de detectar elementos que puedan identificar directa o indirectamente a una fuente confidencial.

Considera:

nombre;

cargo;

institución;

ubicación;

fechas;

descripción de funciones;

fragmentos de mensajes;

documentos;

metadatos;

combinaciones de datos.

Señala cualquier elemento que, incluso sin revelar el nombre, pueda reducir significativamente el grupo de posibles fuentes.

25. Prompt para revisión de lenguaje de certeza

Identifica todas las expresiones del borrador que indiquen certeza, probabilidad, atribución o duda.

Clasifícalas como:

HECHO AFIRMADO

AFIRMACIÓN ATRIBUIDA

INFERENCIA

HIPÓTESIS

INCERTIDUMBRE

Después compara el grado de certeza de la redacción con el grado de evidencia disponible.

Señala cualquier frase que sea más categórica de lo que permiten las fuentes.

26. Prompt para revisar una corrección

Compara:

VERSIÓN PUBLICADA

INFORMACIÓN CORRECTA

Determina:

1. qué afirmaciones resultaron afectadas;
2. si el error alcanza al titular;
3. si modifica la conclusión;
4. qué productos derivados pueden contener el mismo error;
5. qué información interna deberá corregirse para evitar su reproducción.

Propón el alcance factual de la corrección, pero no elimines información necesaria para explicar transparentemente qué cambió.

27. Prompt para identificar contenido potencialmente desactualizado

Revisa este material e identifica afirmaciones que puedan depender del tiempo.

Busca especialmente:

cargos;

autoridades;

normativa vigente;

procesos judiciales;

estadísticas;

precios;

estados de investigaciones;

situaciones institucionales;

enlaces;

fechas futuras o plazos.

Para cada elemento indica:

ÚLTIMA FECHA CONOCIDA

RIESGO DE DESACTUALIZACIÓN

VERIFICACIÓN NECESARIA

28. Prompt para reutilizar información institucional

Cuando se utilice la base de conocimiento institucional deberá evitarse que información histórica se presente automáticamente como actual.

Utiliza la información institucional disponible para identificar antecedentes relacionados con **[ASUNTO]**.

Para cada antecedente indica su fecha.

No asumas que cargos, cifras, situaciones jurídicas, políticas institucionales o circunstancias descritas históricamente continúan vigentes.

Separa:

ANTECEDENTE HISTÓRICO

INFORMACIÓN QUE REQUIERE ACTUALIZACIÓN

INFORMACIÓN ACTUAL VERIFICADA

29. Prompt para elaborar preguntas de entrevista

A partir de la investigación proporcionada, prepara preguntas para entrevistar a **[PERSONA / INSTITUCIÓN]**.

Las preguntas deberán:

estar vinculadas con hechos o documentos concretos;

permitir aclarar contradicciones;

buscar información todavía no disponible;

permitir responder a los señalamientos relevantes;

evitar asumir culpabilidad o hechos todavía no demostrados.

Para cada pregunta indica brevemente qué elemento de la investigación pretende aclarar.

30. Prompt para preparar solicitud de información

A partir del objetivo de investigación **[OBJETIVO]**, identifica qué información debería solicitarse formalmente a **[INSTITUCIÓN]**.

Organiza las solicitudes de forma específica y documental.

Evita preguntas genéricas cuando pueda solicitarse directamente:

documentos; registros; resoluciones; contratos; bases de datos; informes; fechas; comunicaciones; estadísticas; o expedientes.

No asumas que la institución posee información cuya existencia no haya sido comprobada. Cuando corresponda, formula la solicitud de manera que permita determinar si el documento existe.

31. Prompt para revisión final de publicación

Este prompt podrá utilizarse como último control complementario antes de la aprobación humana.

Esta es la versión final propuesta para publicación.

No la reescribas.

Realiza únicamente un control de salida.

Indica **SÍ / NO / REVISAR** para cada elemento:

Los hechos centrales tienen fuente.

Las citas son verificables.

Las cifras tienen respaldo.

Las acusaciones están atribuidas.

El grado de certeza coincide con la evidencia.

No existen hechos nuevos incorporados durante la redacción.

El titular está respaldado.

Las fechas y cargos están actualizados.

No existen datos personales aparentemente innecesarios.

No existe riesgo evidente de identificar una fuente protegida.

Las afirmaciones jurídicas tienen referencia verificable.

Las incertidumbres relevantes permanecen visibles.

Después enumera únicamente los elementos marcados **REVISAR** o **NO**.

No declares que el contenido está “aprobado”. La aprobación corresponde a una persona autorizada.

32. Prompts que deberán evitarse

No deberán utilizarse instrucciones que induzcan a la herramienta a fabricar certeza.

Por ejemplo:

“Demuéstrame que X cometió corrupción.”

“Encuentra pruebas de que esta persona está vinculada con X.”

“Completa los datos que faltan.”

“Escribe una cita probable de lo que quiso decir.”

“Si no encuentras la fuente, utiliza una parecida.”

“Haz que esta acusación suene más contundente.”

“Escribe la nota como si todo estuviera confirmado.”

“Identifica quién probablemente es la fuente anónima.”

El problema de estas instrucciones no radica únicamente en su formulación. Consiste en que orientan al sistema hacia una conclusión determinada o le permiten sustituir evidencia ausente mediante generación.

33. Separación entre prompts de investigación y prompts de redacción

Siempre que resulte razonable, deberán diferenciarse dos momentos.

INVESTIGACIÓN

La IA ayuda a localizar, ordenar, comparar, cuestionar y verificar.

PRODUCCIÓN

La IA trabaja sobre un conjunto de información suficientemente validado para producir un borrador.

No deberá utilizarse un único prompt amplio que simultáneamente:

“investigue todo, saque conclusiones y publique una nota”.

La separación facilita la verificación y reduce el riesgo de que información no comprobada se incorpore directamente al producto.

34. Uso de prompts institucionales compartidos

Cuando Fundamedios identifique una instrucción especialmente eficaz para una actividad recurrente, podrá incorporarla a la biblioteca institucional.

Los prompts institucionales deberán mantenerse dentro de los espacios autorizados y podrán incluir:

plantillas;

variables;

criterios editoriales;

estructuras de análisis;

controles de verificación;

formatos de salida; y

restricciones específicas.

Cuando un prompt constituya parte de una metodología institucional deberá conservarse bajo control de la organización y no exclusivamente en una cuenta personal.

35. Versionamiento de prompts

Los prompts utilizados en procesos recurrentes podrán contar con:

nombre;

versión;

fecha;

finalidad;

herramienta recomendada;

nivel de riesgo;

responsable; y

última revisión.

Ejemplo:

FM-INV-03 / Análisis documental / Versión 2.1 / julio 2026

No será necesario versionar instrucciones cotidianas o exploratorias.

El versionamiento deberá reservarse para prompts que formen parte estable de metodologías institucionales.

36. Evaluación de un prompt antes de institucionalizarlo

Antes de convertir una instrucción en plantilla oficial deberá comprobarse si:

☐ produce resultados consistentes;
☐ diferencia hechos de inferencias;
☐ mantiene atribuciones;
☐ conserva incertidumbres;
☐ permite identificar las fuentes;
☐ no induce a completar información;
☐ funciona adecuadamente con distintos casos;
☐ contiene restricciones suficientes;
☐ no exige introducir información innecesaria;
☐ facilita la revisión humana.

Un prompt no deberá institucionalizarse únicamente porque haya producido un buen resultado en una ocasión.

37. Actualización de la Biblioteca

Esta Biblioteca deberá evolucionar a partir de la experiencia de Fundamedios.

Podrán incorporarse nuevas plantillas cuando aparezcan:

nuevas tareas;

nuevas herramientas;

nuevos formatos editoriales;

errores recurrentes;

necesidades de verificación;

nuevas capacidades de automatización; o

aprendizajes derivados de incidentes.

También deberán retirarse o modificarse prompts que produzcan resultados sistemáticamente deficientes.

38. Regla de oro para prompts institucionales

La calidad de una instrucción no deberá evaluarse únicamente por la calidad aparente del texto que produce.

Un prompt institucional será adecuado cuando contribuya a que el proceso sea **más verificable, controlable y reproducible**.

Por ello, una buena instrucción deberá procurar que la herramienta pueda decir:

“No lo sé.”

“No consta en las fuentes.”

“Existe una contradicción.”

“Esto es una inferencia.”

“Necesita verificación.”

Estas respuestas serán preferibles a una respuesta completa construida sobre información inexistente.

39. Fórmula institucional de cierre

Cuando la naturaleza de una tarea implique riesgo de generación de información no sustentada, podrá incorporarse al final del prompt una instrucción de cierre:

No completes vacíos mediante inferencia. No presentes como hecho aquello que no pueda vincularse con una fuente identificable. Cuando la evidencia sea insuficiente, indícalo expresamente.

Esta fórmula no será obligatoria para todas las interacciones, pero podrá utilizarse como cláusula estándar dentro de los prompts institucionales de investigación y producción de mayor relevancia.

40. Regla final

Los prompts son instrucciones operativas y, por tanto, forman parte del sistema de control de la inteligencia artificial.

Una instrucción deficiente puede producir resultados deficientes incluso cuando la herramienta funcione correctamente.

Sin embargo, ninguna técnica de prompting elimina la posibilidad de error.

Por ello:

Fundamedios utilizará prompts diseñados para favorecer precisión, trazabilidad, atribución y reconocimiento de incertidumbre, pero ningún prompt sustituirá la comprobación de las fuentes ni la decisión humana sobre aquello que puede utilizarse o publicarse.

Anexo 9

Ciclo institucional para el uso de inteligencia artificial

El presente Anexo sintetiza el procedimiento general establecido en el Manual y permite visualizar el ciclo institucional completo de utilización de inteligencia artificial dentro de Fundamedios.

Su finalidad es ofrecer una guía rápida que permita comprender **cómo debe circular la información desde que surge una necesidad de investigación hasta que el producto se publica, se archiva, se corrige o genera aprendizaje institucional**.

El ciclo no deberá entenderse como una secuencia completamente rígida. Dependiendo del tipo de actividad, algunas etapas podrán ser más breves, otras podrán repetirse y determinadas tareas podrán regresar a una fase anterior cuando aparezca nueva información.

La lógica general será:

> **NECESIDAD → PLANIFICACIÓN → RECOLECCIÓN → CLASIFICACIÓN → VERIFICACIÓN → ORGANIZACIÓN → PRODUCCIÓN → REVISIÓN → APROBACIÓN → TRANSPARENCIA → PUBLICACIÓN → ARCHIVO → CORRECCIÓN → APRENDIZAJE**

---

1. Necesidad de información o producto

El ciclo comienza cuando Fundamedios identifica una necesidad concreta.

Puede tratarse, entre otros casos, de:

- investigar un hecho;
- verificar una afirmación;
- preparar una alerta;
- analizar un documento;
- construir una cronología;
- procesar una base de datos;
- preparar una nota;
- producir contenido para redes;
- elaborar un informe;
- revisar una publicación;
- traducir información;
- analizar material audiovisual; o
- realizar otra actividad institucional compatible con el Manual.

Antes de utilizar IA deberá existir una finalidad suficientemente identificable.

Pregunta de control

¿Qué necesitamos hacer y para qué necesitamos inteligencia artificial?

Si la finalidad no puede explicarse razonablemente, deberá definirse antes de continuar.

2. Planificación

Una vez identificada la necesidad, deberá determinarse cómo se utilizará la IA dentro del proceso.

La planificación podrá incluir:

qué se pretende investigar o producir;

qué período interesa;

qué fuentes serán necesarias;

qué información ya se encuentra disponible;

qué información falta;

qué herramienta podría utilizarse;

qué riesgos pueden aparecer; y

qué nivel de verificación será requerido.

En actividades simples esta planificación podrá ser prácticamente inmediata.

En investigaciones complejas deberá existir una definición más clara del objetivo y metodología.

Pregunta de control

¿Qué tarea específica realizará la IA dentro del proceso?

3. Clasificación inicial de la información

Antes de proporcionar información a una herramienta deberá determinarse su nivel de protección.

La clasificación general será:

> **PÚBLICA → INTERNA → CONFIDENCIAL → RESTRINGIDA**

Cuando existan datos personales, fuentes protegidas, datos sensibles o información jurídica reservada deberán aplicarse además las reglas especiales correspondientes.

Decisión

Si la herramienta NO se encuentra autorizada para ese nivel de información → NO CARGAR.

Si puede realizarse la tarea con menos información → MINIMIZAR ANTES DE CARGAR.

Pregunta de control

¿Qué información necesita realmente conocer la herramienta?

4. Selección de la herramienta

La persona usuaria deberá comprobar que la plataforma se encuentre autorizada conforme al Catálogo de Herramientas.

Deberá verificarse:

herramienta correcta;

cuenta institucional;

nivel de información autorizado;

permisos;

memoria;

integraciones;

funcionalidades; y

restricciones específicas.

Decisión

> **HERRAMIENTA AUTORIZADA + FINALIDAD AUTORIZADA + INFORMACIÓN AUTORIZADA = CONTINUAR**

Si alguno de estos elementos no coincide, deberá utilizarse otra alternativa o solicitarse evaluación.

5. Minimización y preparación

Antes de cargar información deberá retirarse aquello que no sea necesario.

Dependiendo del caso podrán eliminarse:

- nombres;
- identificadores;
- datos de contacto;
- metadatos;
- anexos;
- páginas irrelevantes;
- credenciales;
- información de fuentes;
- datos sensibles;
- información descartada.

Cuando sea posible deberá utilizarse una versión preparada específicamente para procesamiento mediante IA.

Pregunta de control

¿La herramienta puede cumplir la tarea recibiendo menos información?

6. Recolección o procesamiento inicial

La IA podrá utilizarse entonces para:

BUSCAR — LOCALIZAR — EXTRAER — CLASIFICAR — RESUMIR — TRANSCRIBIR — COMPARAR — ORGANIZAR

Los resultados de esta etapa deberán considerarse preliminares.

La herramienta podrá descubrir información útil, pero todavía no deberá asumirse que esa información se encuentra validada.

Resultado esperado

Un conjunto de:

fuentes posibles; hallazgos; documentos; referencias; datos; contradicciones; preguntas nuevas; e información pendiente.

---

7. Registro de procedencia

Todo hallazgo relevante deberá conservar una relación con su origen.

La cadena básica será:

DATO → FUENTE → UBICACIÓN

Cuando corresponda también deberá registrarse:

fecha; autor; enlace; página; documento; estado de verificación.

Un dato que no pueda vincularse con una fuente deberá permanecer como pendiente.

---

8. Verificación

La información recopilada deberá someterse al nivel de verificación que corresponda.

La verificación deberá comprobar, según el caso:

existencia de la fuente;

correspondencia entre fuente y afirmación;

autoría;

fecha;

contexto;

vigencia;

cifras;

citas;

identidades;

contradicciones; y

calidad de la evidencia.

Estados posibles

VERIFICADA

CONDICIONADA

PENDIENTE

CONTRADICTORIA

DESCARTADA

La IA no decidirá por sí sola cuál de estos estados corresponde.

9. Punto de retorno: información nueva

Si durante la verificación aparece una nueva fuente, afirmación o línea de investigación, esta deberá regresar a la fase de recolección.

El ciclo será:

> **NUEVO HALLAZGO → NUEVA BÚSQUEDA → NUEVA VERIFICACIÓN**

No deberá incorporarse directamente a producción únicamente porque aparezca durante una etapa avanzada.

10. Organización de la información validada

Antes de producir un contenido deberá organizarse el material.

Deberán distinguirse:

hechos;

fuentes;

citas;

versiones;

contexto;

inferencias;

incertidumbres;

información descartada;

datos restringidos.

La información deberá almacenarse dentro del repositorio institucional correspondiente.

Resultado esperado

BASE DE INFORMACIÓN DEPURADA Y LISTA PARA PRODUCCIÓN

11. Preparación del input de producción

No deberá enviarse automáticamente todo el expediente a la herramienta encargada de redactar.

Deberá prepararse un conjunto específico que incluya únicamente lo necesario.

Una estructura posible será:

HECHOS VERIFICADOS

CITAS VERIFICADAS

POSICIONES DE LAS PARTES

CONTEXTO

INFORMACIÓN CONDICIONADA

RESTRICCIONES

Regla

El expediente puede ser amplio; el input debe ser selectivo.

12. Construcción del prompt

La instrucción deberá definir:

qué producir;

para quién;

con qué fuentes;

en qué formato;

con qué tono;

qué no debe hacer;

qué incertidumbres debe mantener;

qué información no puede completar.

En productos relevantes deberá incorporarse expresamente la prohibición de generar hechos faltantes.

13. Producción asistida por IA

La herramienta podrá:

estructurar;

resumir;

redactar;

reformular;

traducir;

adaptar;

proponer titulares;

generar tablas;

preparar contenidos derivados.

Todo resultado generado será considerado:

BORRADOR

hasta completar la revisión correspondiente.

14. Primera revisión humana

La persona que utilizó la IA deberá revisar íntegramente el resultado.

Deberá comprobar:

fidelidad;

ausencia de hechos nuevos;

atribuciones;

citas;

cifras;

contexto;

incertidumbres;

protección de información;

coherencia.

Pregunta de control

¿La IA escribió únicamente aquello que Fundamedios puede sostener?

15. Punto de retorno: hecho nuevo generado durante la redacción

Si la IA incorpora un dato no incluido en los insumos:

DETENER → IDENTIFICAR FUENTE → INVESTIGAR → VERIFICAR

No deberá verificarse preguntando simplemente a la misma herramienta si está segura.

La producción no podrá convertirse en una vía indirecta para incorporar información no investigada.

16. Revisión editorial

Cuando corresponda, una segunda persona deberá revisar el producto.

El nivel dependerá del riesgo:

> **ORDINARIO → REFORZADO → ESPECIAL**

La revisión deberá concentrarse especialmente en las afirmaciones centrales.

17. Revisión especializada

Cuando el producto lo requiera, deberá intervenir:

JURÍDICA

PROTECCIÓN DE DATOS

SEGURIDAD

TÉCNICA O TEMÁTICA

La revisión especializada deberá realizarse antes de la aprobación final.

18. Revisión adversarial

En productos de riesgo medio o alto podrá utilizarse IA nuevamente para buscar errores.

Su función será cuestionar el producto:

¿Qué no está respaldado?

¿Qué contradice una fuente?

¿Qué falta?

¿Dónde hay exceso de certeza?

¿Qué puede afectar derechos?

La revisión adversarial será complementaria.

LA IA PUEDE ALERTAR; NO PUEDE APROBAR.

19. Control final

Antes de publicar deberá comprobarse:

HECHOS

¿Están verificados?

FUENTES

¿Son identificables?

CITAS

¿Son reales?

CIFRAS

¿Son correctas?

ATRIBUCIONES

¿Se mantienen?

CONTEXTO

¿Es suficiente?

CONTRASTE

¿Se realizó cuando correspondía?

DATOS PERSONALES

¿Son necesarios?

FUENTES PROTEGIDAS

¿Continúan protegidas?

TITULAR

¿Está respaldado?

DISCLOSURE

¿Corresponde incorporarlo?

APROBACIÓN

¿La versión exacta fue autorizada?

20. Aprobación humana

La decisión final de publicación corresponderá a una persona autorizada.

El flujo deberá detenerse si existe una duda material que no pueda resolverse.

Regla

> **IA PROPONE / HUMANO DECIDE**

La aprobación deberá recaer sobre la versión exacta que posteriormente será publicada.

21. Transparencia

Antes de publicación deberá evaluarse si resulta necesario informar al público sobre el uso de inteligencia artificial.

La pregunta principal será:

¿Conocer la intervención de IA cambiaría razonablemente la forma en que el público interpreta el contenido?

Si la respuesta es sí, deberá realizarse una divulgación proporcional.

Cuando exista material sintético susceptible de confundirse con evidencia real, la identificación deberá ser clara.

22. Publicación

Una vez aprobada, la versión podrá publicarse.

La automatización podrá ejecutar técnicamente una publicación previamente autorizada.

No deberá automatizarse la decisión editorial de publicar cuando el Manual exija intervención humana.

Regla

> **VERSIÓN APROBADA = VERSIÓN PUBLICADA**

23. Productos derivados

Si posteriormente la publicación se transforma en:

- redes sociales;
- boletín;
- traducción;
- video;
- gráfico;
- presentación;
- resumen;
- nueva alerta;

cada adaptación deberá conservar la precisión del producto original.

Una nueva generación mediante IA no deberá introducir hechos nuevos.

24. Archivo y trazabilidad

Después de publicación deberán conservarse, según el nivel de riesgo:

fuentes centrales;

documentos originales;

verificaciones;

versiones relevantes;

prompts sustantivos;

outputs cuando resulte necesario;

aprobaciones;

versión publicada.

No será necesario conservar toda interacción insignificante.

Objetivo

PODER RECONSTRUIR LO IMPORTANTE

25. Monitoreo posterior

Después de publicación podrán aparecer:

nueva información;

respuestas;

reclamos;

correcciones;

cambios jurídicos;

cambios de contexto;

errores detectados.

Cuando esto ocurra deberá determinarse si corresponde actualizar, corregir o revisar el producto.

26. Corrección

Si existe un error:

> **IDENTIFICAR → CONTENER → CORREGIR → RASTREAR ORIGEN → ACTUALIZAR SISTEMAS**

La corrección deberá alcanzar, cuando corresponda:

publicación principal;

redes sociales;

traducciones;

productos derivados;

bases de conocimiento;

memorias;

repositorios internos.

La corrección pública no deberá dejar activo internamente el dato que produjo el error.

27. Gestión de incidentes

Cuando exista un incidente relevante deberá activarse el procedimiento correspondiente.

El flujo será:

> **DETECTAR → REPORTAR → CONTENER → PRESERVAR EVIDENCIA → EVALUAR → CORREGIR → RECUPERAR**

Cuando intervengan datos personales, fuentes protegidas o seguridad deberá aplicarse además el procedimiento específico correspondiente.

28. Aprendizaje institucional

Después de un incidente, corrección o revisión importante deberá determinarse si es necesario modificar:

una herramienta;

un prompt;

una configuración;

un permiso;

una clasificación de riesgo;

una capacitación;

un procedimiento;

un anexo;

el propio Manual.

Resultado

EL ERROR NO SOLO SE CORRIGE: TAMBIÉN DEBE PRODUCIR APRENDIZAJE

29. Memoria institucional

Los aprendizajes, metodologías, prompts y materiales institucionales que tengan valor permanente podrán incorporarse, de manera selectiva, al conocimiento de Fundamedios.

No deberá incorporarse indiscriminadamente:

información falsa;

borradores;

datos innecesarios;

información restringida;

fuentes protegidas;

material desactualizado.

La memoria deberá ser gestionada y depurada.

30. Representación resumida del ciclo

El ciclo completo podrá visualizarse de la siguiente manera:

1. NECESIDAD

↓

2. PLANIFICACIÓN

↓

3. CLASIFICACIÓN DE INFORMACIÓN

↓

4. SELECCIÓN DE HERRAMIENTA

↓

5. MINIMIZACIÓN

↓

6. RECOLECCIÓN / PROCESAMIENTO

↓

7. REGISTRO DE FUENTES

↓

8. VERIFICACIÓN

↓

9. ORGANIZACIÓN

↓

10. INPUT DE PRODUCCIÓN

↓

11. GENERACIÓN

↓

12. REVISIÓN HUMANA

↓

13. REVISIÓN ESPECIALIZADA CUANDO CORRESPONDA

↓

14. APROBACIÓN

↓

15. DISCLOSURE CUANDO CORRESPONDA

↓

16. PUBLICACIÓN

↓

17. ARCHIVO Y TRAZABILIDAD

↓

18. MONITOREO

↓

19. CORRECCIÓN / INCIDENTE

↓

20. APRENDIZAJE INSTITUCIONAL

↓

NUEVO CICLO

31. Puntos de retorno

El sistema no será completamente lineal.

Existirán al menos cuatro puntos principales de retorno:

NUEVA INFORMACIÓN DURANTE VERIFICACIÓN

Regresa a **RECOLECCIÓN**.

INFORMACIÓN NUEVA DURANTE REDACCIÓN

Regresa a **INVESTIGACIÓN Y VERIFICACIÓN**.

ERROR DURANTE REVISIÓN

Regresa a **PRODUCCIÓN o VERIFICACIÓN**, según su origen.

ERROR DESPUÉS DE PUBLICACIÓN

Regresa a **FUENTE + VERIFICACIÓN + SISTEMAS INTERNOS**, además de realizar la corrección pública.

Estos retornos forman parte normal del sistema y no deberán considerarse excepciones.

32. Semáforo operativo

Como orientación rápida podrá utilizarse el siguiente criterio:

🟢 **CONTINUAR**

La información está clasificada, la herramienta está autorizada y los controles necesarios están cumplidos.

🟡 **VERIFICAR / CONSULTAR**

Existe duda sobre una fuente, dato, clasificación, permiso, riesgo o interpretación.

🔴 **DETENER**

Existe información restringida en una herramienta no autorizada, una fuente puede quedar expuesta, falta una verificación indispensable, se pretende publicar sin aprobación o existe cualquier riesgo crítico no controlado.

El semáforo no sustituye las matrices del Manual, pero puede facilitar decisiones cotidianas.

33. Las siete preguntas del ciclo

Cuando una persona necesite una versión todavía más sencilla del procedimiento, deberá poder responder siete preguntas:

1. ¿Qué estoy tratando de hacer?

2. ¿Qué información voy a proporcionar?

3. ¿La herramienta puede recibirla?

4. ¿De dónde proviene lo que me está diciendo la IA?

5. ¿Ya verifiqué aquello que voy a utilizar?

6. ¿Quién revisará y aprobará el resultado?

7. ¿Podremos reconstruir posteriormente cómo llegamos a la publicación?

Si alguna respuesta esencial es negativa, deberá resolverse antes de avanzar.

34. Principio de control progresivo

Los controles deberán aumentar a medida que la información se acerca al público o a una consecuencia externa.

Una búsqueda preliminar tendrá controles menores.

Una información utilizada como fuente tendrá controles mayores.

Una acusación incorporada a una publicación tendrá controles todavía superiores.

Una acción automatizada que pueda producir efectos externos requerirá control específico.

Esto puede expresarse así:

> **EXPLORACIÓN < ANÁLISIS < PRODUCCIÓN < PUBLICACIÓN < ACCIÓN EXTERNA**

A mayor impacto, mayor supervisión.

35. Responsabilidad a lo largo del ciclo

No existirá una única persona responsable de todo el proceso en todos los casos.

La responsabilidad seguirá la función:

QUIEN INVESTIGA → responde por los controles de investigación.

QUIEN VERIFICA → responde por la verificación realizada.

QUIEN PRODUCE → responde por revisar el resultado generado.

QUIEN REVISA → responde por el alcance de su revisión.

QUIEN APRUEBA → responde por ejercer efectivamente la aprobación.

QUIEN ADMINISTRA → responde por los controles tecnológicos bajo su competencia.

La existencia de varias funciones no deberá diluir la responsabilidad ni concentrarla automáticamente en una sola persona.

36. Objetivo final del ciclo

El procedimiento completo persigue que Fundamedios pueda utilizar inteligencia artificial de forma intensiva cuando resulte útil sin perder control sobre cuatro elementos esenciales:

LA INFORMACIÓN

LA EVIDENCIA

LA DECISIÓN

LA RESPONSABILIDAD

La tecnología podrá cambiar.

Las herramientas utilizadas podrán ser sustituidas.

Las capacidades de los modelos podrán aumentar.

El ciclo institucional deberá continuar asegurando que la información que llega al público pueda ser explicada, verificada y respaldada por Fundamedios.

37. Regla final del Anexo

El sistema podrá resumirse en una sola idea:

La inteligencia artificial puede intervenir en todas las etapas del trabajo, pero ninguna etapa deberá perder su control institucional por el hecho de haber sido automatizada.

El ciclo terminará únicamente de manera provisional con la publicación.

La conservación, revisión, corrección y aprendizaje devolverán continuamente información al sistema y permitirán que el modelo institucional de uso de IA evolucione con la experiencia.



La IA apoya,
la responsabilidad es humana.

FUNDAMEDIOS · ECUADOR Chequea PERIODISMO CON RIGOR · Liberties 360°

José Padilla N 330 e Iñaquito, Edificio Platinum, piso 10, oficina 1002 · Quito, Ecuador · (593) 2 601 9956 · info@fundamedios.org · www.fundamedios.org

Ilustraciones e íconos generados con IA.